Two ICS Security Datasets and Anomaly Detection Contest on the HIL-based Augmented ICS Testbed

Hyeok-Ki Shin*
The Affiliated Institute of ETRI
Daejeon, Republic of Korea
hkshin721@nsr.re.kr

Woomyo Lee*
The Affiliated Institute of ETRI
Daejeon, Republic of Korea
wmlee@nsr.re.kr

Jeong-Han Yun*
The Affiliated Institute of ETRI
Daejeon, Republic of Korea
dolgam@nsr.re.kr

Byung Gil Min
The Affiliated Institute of ETRI
Daejeon, Republic of Korea
bgmin@nsr.re.kr

ABSTRACT

Security datasets with various operating characteristics and abnormal situations of industrial control system (ICS) are essential to develop artificial intelligence (AI)-based control system security technology. In this study, we built a hardware-in-the-loop (HIL)-based augmented ICS (HAI) testbed and developed ICS security datasets. Here, we introduce the second dataset (HAI 21.03), which was developed with the user feedback of the first released version (HAI 20.07). All HAI datasets are publicly available at https://github.com/icsdataset/hai. HAI 21.03 was expanded by adding data points and normal/attack scenarios to HAI 20.07. We also held an AI-based anomaly detection contest (HAICon 2020) utilizing the HAI datasets developed so far, giving many AI researchers an opportunity to discuss and share ideas for ICS anomaly detection research. This paper presents the results of the HAICon 2020. The results of the top teams in the competition can be used as a performance comparison criterion when using HAI 21.03.

CCS CONCEPTS

• **Information systems → Process control systems**; • **Computing methodologies → Anomaly detection**.

KEYWORDS

security dataset, industrial control system, testbed, hardware-in-the-loop, anomaly detection, artificial intelligence

ACM Reference Format:
Hyeok-Ki Shin, Woomyo Lee, Jeong-Han Yun, and Byung Gil Min. 2021. Two ICS Security Datasets and Anomaly Detection Contest on the HIL-based Augmented ICS Testbed. In *Cyber Security Experimentation and Test Workshop (CSET '21), August 9, 2021, Virtual, CA, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3474718.3474719

*These authors contributed equally to this research.

CSET '21, August 9, 2021, Virtual, CA, USA

ACM ISBN 978-1-4503-9065-1/21/08…$15.00
https://doi.org/10.1145/3474718.3474719

1 INTRODUCTION

The detection of possible cyber attacks or unexpected failures in industrial control systems (ICSs), such as water pumps, power grids, and nuclear power plants, is crucial for the prevention of dire consequences [2]. Although there is a growing number of studies on ICS security, there is still a lack of open datasets that can be used for research. Thus, our goal is to create datasets for ICS security researchers working on anomaly detection. For this purpose, we first implemented a hardware-in-the-loop (HIL)-based augmented ICS (HAI) testbed[3] to generate accurate datasets for various scenarios while minimizing human effort.

Table 1: Release overview of HAI security datasets. HAI 20.07 is a bug fix release of the first version [4] and the 2nd version HAI 21.03 is released in March 2021.

Version	Data points (points/sec)	Training set			Test set			
		File (CSV)	Duration (hours)	Size (MB)	File (CSV)	Attack count	Duration (hours)	Size (MB)
HAI 21.03	78	train1	60	110	test1	5	12	22
		train2	63	116	test2	20	33	62
		train3	229	246	test3	8	30	56
					test4	5	11	20
					test5	12	26	48
HAI 20.07	59	train1	86	127	test1	28	81	119
		train2	91	98	test2	10	42	62

An HAI dataset (HAI 20.07[1]) [4] was released at https://github.com/icsdataset/hai. After the first release of this dataset, we developed a new version of the dataset (HAI 21.03) using the HAI testbed. Considering user opinions on the first dataset, we focus on three key issues for ICS anomaly detection research.

- Reconfiguration of the testbed: The scaling and biasing factors of the analog signal between the HIL simulator and the physical system were reconfigured to increase the mutual influence, hence the establishment of a new testbed with new response characteristics.
- Causality of the dataset: Data collection points were added to clearly interpret the causal relationship of the control process. (e.g., set points, process variables, and control outputs)
- Various normal and attack scenarios: some scenarios were additionally developed in the reconfigured testbed for learning and anomaly detection performance evaluation.

[1]The initial version name of the HAI dataset was HAI 1.0 [4], but the version numbering scheme was changed to specify the release date of the dataset.

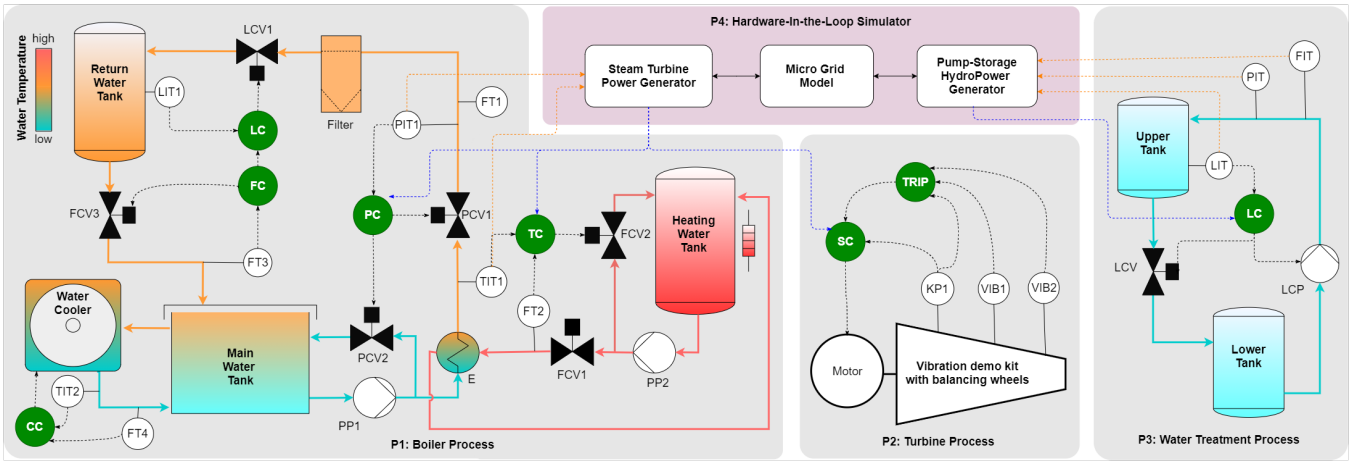

Figure 1: Process control architecture of HAI Testbed

Based on the HAI datasets, we held an artificial intelligence (AI)-based anomaly detection contest (HAICon 2020). The best detection results of the contest can be used as a performance comparison criterion for researchers using HAI datasets. In this paper, we introduce a new HAI dataset and the results of the AI contest.

The remainder of this paper is organized as follows: Section 2 introduces the changes in the HAI testbed [3] to generate a new dataset that differs from HAI 20.07 [4]. Section 3 describes the second HAI dataset (HAI 21.03). Section 4 presents the results of HAICon 2020. Finally, Section 5 presents the conclusions of this study and future work.

2 RECONFIGURATION OF THE HAI TESTBED

The HAI testbed designed to easily replicate various operating environments, consisted of a boiler, turbine, water-treatment component, and HIL simulator [3].

As shown in Figure 1, the process architecture of the HAI testbed consists of four primary processes: the boiler process (P1), turbine process (P2), water-treatment process (P3), and HIL simulator (P4). The HIL simulation enhances the correlation between the three real-world processes at the signal level by simulating thermal power generation and pumped-storage hydropower generation scenarios. The boiler and turbine processes were used to simulate the thermal power plant, and the water treatment process was used to simulate the pumped-storage hydropower plant.

A new HAI testbed with new characteristics was developed through the reconfiguration of several sets of system parameters. The first set of parameters is the scaling and biasing factors of the HIL analog input/output signals. These factors control the strength of the mutual influence between virtual and physical systems. Increasing this intensity makes the causal relationship between the virtual and physical systems clearer. As a result, the variation width of the HIL analog input/output increases, and eventually, a greater influence is transmitted to the physical system. The setting values were readjusted for five analog inputs and three analog outputs of the HIL simulator. It was helpful for the fidelity of the dataset because it is possible to combine various and complex processes, but there is a limit to creating a perfect real world because some functions have been expanded through the virtual system.

The other sets are the gains of the proportional–integral–derivative (PID) controller that is used in most ICSs to regulate the temperature, pressure, flow, level, and other industrial process variables. In the HAI testbed, six PID controllers are operated in three distributed control systems (DCSs), and two PID controllers are used for valve control in the HIL simulation. These gains are experimentally changed to new values for system stabilization following the reconfiguration of the HIL analog input/output settings. As a result, the change pattern of the process variable appears differently during the process regulation.

3 HAI SECURITY DATASET v 21.03

We developed a second security dataset, HAI 21.03, in the new HAI testbed environment. The training data were developed based on more abundant normal scenarios, and the test data were obtained through the realization of more diverse attack scenarios to enable the evaluation of detection performance in various cases.

3.1 Process historian

HAI 20.07 was collected from 59 data points for five process controllers (boiler pressure control, boiler water level control, boiler flow control, turbine speed control, and water-treatment level control). In HAI 21.03, 78 data points were collected in six process controllers with added turbine trip control. In addition, data collection points were added to configure the dataset of the inputs (setpoint (SP), process variables (PV)) and outputs (Contol variable (CV)) of the controllers. Accordingly, the causal relationship between control inputs and outputs can be clearly analyzed.

3.2 Training set

The operator is assumed to operate the control facility in a routine manner via the human–machine interface (HMI), and the simulator variables associated with power generation in the HIL simulator are changed. The operator monitors the PV values given by the

Table 2: Normal operations of HAI 21.03: schedules of set points during one day

No	Set points								Start time	
	Pressure		Level		Flow rate		Temperature			
1	0.1	(±0.002)	440	(±9)	1,100	(±22)	32	(0)	03:00	(±10)
2	0.03	(±0.001)	400	(±8)	1,100	(±22)	32	(0)	04:30	(±10)
3	0.1	(±0.002)	400	(±8)	1,100	(±22)	32	(±1)	06:00	(±10)
4	0.1	(±0.002)	400	(±8)	900	(±18)	32	(0)	08:30	(±10)
5	0.1	(±0.002)	380	(±8)	1,100	(±22)	32	(0)	10:00	(±10)
6	0.06	(±0.001)	420	(±8)	1,000	(±20)	32	(0)	12:00	(0)
7	0.1	(±0.002)	400	(±40)	1,100	(±22)	32	(0)	14:30	(±10)
8	0.1	(±0.002)	400	(±8)	1,000	(±60)	33	(±1)	17:00	(±10)
9	0.1	(±0.002)	400	(±8)	1,100	(±22)	32	(±1)	19:30	(±10)
10	0.1	(±0.002)	400	(±8)	1,100	(±22)	32	(±1)	22:00	(±10)

current sensor displayed on the HMI and adjusts the SPs of the various control devices to operate the system.

An HMI operation task scheduler was used to periodically set the SPs and HIL simulator variables to random or predefined values within the normal range in which the entire process stably operates.

HAI 21.03 increased the number of benign scenarios and collection period as compared to HAI 20.07. Normal (learning) data were collected for 11 days, in which 10 benign scenarios were implemented, as shown in Table 2.

3.3 Test set

We attempted to develop abnormal data for various security threat situations in which an attacker takes control of the HMI, PLC, and DCS and causes malfunction.

Attack scenarios are based on the PCL (process control loop) model shown in Figure 2 and implemented by an automation tool that executes attacks predefined by the user remotely through OPC communication. This tool can implement stealth attacks and manipulate multiple target points sequentially or in parallel. Metadata for all attacks performed through the tool (e.g., target of attack and time of attack) are stored and used for data labeling.

All the attack scenarios were configured based on the four variables of the feedback control loop, namely, SP, PV, CV, and control parameter (CP). Abnormal behaviors occurred when some of the parameters were not within the limits of the normal range or were in unexpected states due to attacks, malfunctions, and failures.

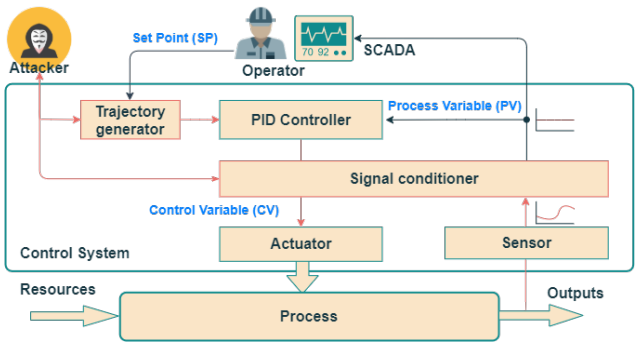

Figure 2: Attack model based on a process control loop

Since 2019 [3, 4], attack scenarios have been continuously developed and implemented by considering the attack target, attack time, and method for each feedback control loop. HAI 20.07 collected abnormal datasets that implemented 14 single-attack scenarios and 19 complex-attack scenarios over five days. HAI 21.03 performed 25 single-attack scenarios and 25 combined-attack scenarios for five days. In HAI 21.03, attack scenarios for short-term (ST) attacks, control command forgery attacks (P1-FC and P2-SC), and safety instrumentation systems (SC-TC) were also implemented. The details of the abnormal data used in HAI 21.03 are given in Appendix A.

3.4 Data labeling

HAI 21.03 includes three CSV files as training data and five CSV files as test data. The time-series data in each CSV file satisfy time continuity and include 84 columns. The first column represents the observed time as "yyyy-MM-dd hh:mm:ss," whereas the next 78 columns provide the recorded SCADA data points. The last four columns provide data labels for whether an attack occurred or not, where the attack column was applicable to all processes, and the other three columns were for the corresponding control processes.

4 HAICon 2020: AI-BASED ICS ANOMALY DETECTION CONTEST ON HAI DATASETS

In order to revitalize research, discover ideas, and improve HAI dataset through participants' feedback, we held an AI contest "HAICon 2020", hosted by the Affiliated Institute of ETRI and supported by the Korea Institute of Information Security and Cryptology (KIISC).

4.1 Procedures and rules

HAICon 2020 competes for the anomaly detection performance of a semi-supervised learning model that detects abnormal (or unknown) behaviors that do not appear in the training data under normal conditions. Participants were provided with datasets(raw version of HAI 21.03), baseline model, and evaluation tool eTaPR (enhanced version of TaPR [1]) to help them understand how the competition was conducted. The outline of the competition procedures and rules is as follows:

- Participation: Applicants can apply for participation individually or as a team through an online competition website[2]. A team can comprise up to five people.
- Data usage: Do not allow the use of external data.
- Model development and evaluation: The model can be trained using the training data, and the trained model can be verified using the validation data and eTaPR evaluation tool. This process refers to the baseline model provided on the competition website.
- Result submission: The detection results for the test data are submitted online in the specified CSV format. It can be submitted up to three times a day, and only individuals or team leaders can submit the results.
- Monetary prizes: The total prize money is 20 million KRW: 1st place = 10 million, 2nd place = 5 million, 3rd place = 3 million, and 4th place (two teams) = 1 million each.

[2]https://dacon.io/competitions/official/235624/overview/description

4.1.1 Dataset preparation. The competition dataset was divided into three types of data: training, validation, and test. These data were released to participants through the competition website on the day of the competition.

- *Training set* is a set of data collected during normal operations. They are provided in three files with time continuity because they were collected in three different terms.
- *Validation set* includes five out of 50 attacks and includes the attack label. Participants roughly check the performance of the currently trained model.
- *Test set* was collected in 45 out of 50 attacks. They were provided separately in four files with temporal continuity and do not contain an attack label. Participants must submit the detection results for these data.

Accordingly, the competition dataset was de-identified by shuffling the columns of all data and renaming the columns ("C01" to "C79"). This is because it is generally difficult to obtain information for the identification of data points in ICSs. This step also aims to prevent participants from improving the detection performance of test data for competitions using information from HAI 20.07, which has already been disclosed.

4.1.2 Evaluation metric: eTaPR. The competition performance metric, eTaPR, was provided at the competition website in the form of a Python wheel package for an accurate performance evaluation of anomaly detection for time-series data.

Because the F1 score, which is used most often, is scored according to the accumulation of detected time, it is difficult to properly evaluate the results of detecting attacks that are high in risk and occur over a short period of time. Therefore, eTaPR reflects the number of attack occurrences in the evaluation factor to enable partial anomaly detection.

To properly use eTaPR, four parameters must be set in consideration of the dynamic characteristics of the dataset. In the eTaPR Python package, we set the parameter values to $\alpha = 0.5$, $\rho = 0.1$, $\pi = 0.7$, and $\delta = 180$ in consideration of the characteristics of the competition dataset. These values are not allowed to change.

4.1.3 Baseline detection models. We provided baseline models (source codes to detect anomalies) for the HAI datasets on the competition website. Baselines help participants understand how a competition is conducted using HAI datasets. It contains examples of data preprocessing, model training, anomaly detection, and performance evaluation using the competition dataset.

First, data pre-processing to remove high-frequency noise was performed on the training data, and then RNN-based prediction model was trained. If the difference between the predicted results through the prediction model and the actual value exceeded a predetermined threshold, then it was judged as an abnormal situation.

Next, the detection performance of the baseline model was verified using the verification data. The detected results could be checked using the provided attack label and eTaPR library.

Finally, the detection results of the baseline model for the test data were written to the CSV file of the competition standard.

4.1.4 Public/private scoring. When participants submitted a detection result file through the competition website, public and private eTaPR scores were automatically calculated. The public score is an eTaPR score of the submitted result for 30% of the test data. This score and ranking (see Figure B1) were only updated in real time on the leaderboard to prevent overfitting of the test data. The private score is the eTaPR score for the entire detection result submitted by the participant. It was not revealed on the website before the end of the competition, and the winner was determined based on this score.

4.2 Results

Most of the 890 participating teams were Korean students and researchers. According to the rules of competition, some foreigners teamed up with Koreans to participate. 311 teams submitted the detection results and placed their names on the leaderboard.

The final scores of the top teams are shown in Table B1. The final ranking was determined by comprehensively reviewing the model reproduction results through code submission and the private score.

Meanwhile, the baseline model ranks 156th, but the top 7 teams submitted anomaly detection models developed based on the baseline model.It would have been more advantageous to improve the performance through baseline model optimization rather than developing a new model for a short competition period of approximately two months.

Obviously, even a very simple model based on RNN achieved good performance. It seems that data pre-processing and post-processing of prediction errors have a great influence on the detection performance improvement rather than the learning model. Also the method of ensemble using multiple time windows also helps to improve performance.

5 CONCLUSIONS AND FUTURE WORK

Our ultimate goal is to create a generalized evaluation framework that allows AI-based anomaly detection systems to detect and evaluate various types of attacks under various operating conditions.

We released the second dataset, HAI 21.03, through an online anomaly detection contest, HAICon 2020, which was held to facilitate ICS security research.

Currently, we are preparing a third dataset and a second HAICon, HAICon 2021. To diversify the attack impact, we improved the control logic relationship between the HIL and physical systems and attached a cooling device to the boiler system. In addition, we are studying how to create various normal/attack data to verify the performance of ICS security technologies according to site requirements.

REFERENCES

[1] Won-Seok Hwang, Jeong-Han Yun, Jonguk Kim, and Hyoung Chun Kim. 2019. Time-Series Aware Precision and Recall for Anomaly Detection: Considering Variety of Detection Result and Addressing Ambiguous Labeling. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. ACM, 2241–2244.

[2] Siddhartha Kumar Khaitan and James D McCalley. 2014. Design techniques and applications of cyberphysical systems: A survey. In *IEEE Systems Journal*.

[3] Hyeok-Ki Shin, Woomyo Lee, Jeong-Han Yun, and HyoungChun Kim. 2019. Implementation of Programmable CPS Testbed for Anomaly Detection. In *12th USENIX Workshop on Cyber Security Experimentation and Test(CSET 19)*.

[4] H.-K. Shin, W. Lee, J.-H. Yun, and H. C. Kim. 2020. HAI 1.0: HIL-based Augmented ICS Security Dataset. In *13th USENIX Workshop on Cyber Security Experimentation and Test(CSET 20)*.