HAI SECURITY DATASET

HIL-BASED AUGMENTED ICS (HAI) SECURITY DATASET WAS COLLECTED FROM A REALISTIC INDUSTRIAL CONTROL SYSTEM (ICS) TESTBED AUGMENTED WITH A HARDWARE-IN-THE-LOOP (HIL) SIMULATOR THAT EMULATES STEAM-TURBINE POWER GENERATION AND PUMPED-STORAGE HYDROPOWER GENERATIONON

Document version: 3.0
Release date: April 2022

RELEASE HISTORY

HAI is a security dataset that includes both the normal and abnormal behaviors for ICS anomaly detection research. The normal dataset was collected continuously for several days. Moreover, the abnormal dataset was collected based on various attack scenarios with the six control loops in three different types of industrial control devices, namely the Emerson Ovation, GE Mark-VIe, and Siemens S7-1500. Here, a control loop refers to a system comprising all the software functions required to measure and adjust the variable that controls a process.

Version History

Two major versions of HAI datasets have been released until now. Each dataset consists of several CSV files, and each file satisfies time continuity. The quantitative summary of each version are as follows:

Release Version	Points (per sec)	Normal Dataset			Abnormal Dataset			
		Files	Interval (hours)	Size (MB)	Files	Attack Counts	Interval (hours)	Size (MB)
HAI 22.04	86	train1.csv	26	50.7	test1.csv	7	24	48.2
		train2.csv	56	108.9	test2.csv	17	23	44.5
		train3.csv	35	66.7	test3.csv	10	17.3	33.4
		train4.csv	24	45.7	test4.csv	24	36	69.5
		train5.csv	66	125.6	-			
		train6.csv	72	136.8				
		SUM	**279**	**534.4**	**SUM**	**58**	**100.3**	**195.6**
HAI 21.03	78	train1.csv	60	110	test1.csv	5	12	22
		train2.csv	63	116	test2.csv	20	33	61
		train3.csv	229	245	test3.csv	8	30	55
		-			test4.csv	5	11	20
					test5.csv	12	26	47
		SUM	**352**	**471**	**SUM**	**50**	**112**	**205**
HAI 20.07	59	train1.csv	86	127	test1.csv	28	81	119
		train2.csv	91	98	test2.csv	10	42	62
		SUM	**177**	**225**	**SUM**	**38**	**123**	**181**

Note: The version numbering follows a date-based scheme, where the version number indicates the released date of a HAI dataset. HAI 20.07 is the bug-fixed version of HAI v1.0 released in February 2020.

Document Change Logs

Version	Release Date	Changes	Page(s)
v3.0	Apr. 29, 2022	***Major revision for HAI 22.04***	
		+ Version history for HAI 22.04	*01*
		+ Brief description of the boiler cooling system	*03-05*
		+ Detailed description of the boiler cooling controller	*08*
		+ 8 more data points	*10 – 13*
		+ 12 more attack scenarios	*15 – 17*
		+ Correct some errors on the attack scenarios	
		+ Details of HAI 22.04	*18 – 21*
		+ Citing datasets	*28*
v2.0	Feb. 17, 2021	***Major revision for HAI 21.03***	
		+ Brief description of the turbine trip control	*09*
		+ 19 more data points	*10 – 13*
		+ 11 more attack scenarios	*15 – 17*
		- Description related to multiple attacks	*15*
		+ Details of HAI 21.03	*22 – 24*
		+ Changes to HAI 20.07	*25 – 27*
v1.1	Jul. 22, 2020	***Minor revision for HAI 20.07***	
		+ New version numbering scheme	*All*
		+ Value ranges and description of data points	*10 – 13*
		+ Time duration in attack timetable	*25 – 27*
v1.0	Feb. 17, 2020	***Initial release for HAI v1.0 (20.02)***	*All*

HAI SECURITY DATASET

HIL-BASED AUGMENTED ICS (HAI) SECURITY DATASET WAS COLLECTED FROM A REALISTIC INDUSTRIAL CONTROL SYSTEM (ICS) TESTBED AUGMENTED WITH A HARDWARE-IN-THE-LOOP (HIL) SIMULATOR THAT EMULATES STEAM-TURBINE POWER GENERATION AND PUMPED-STORAGE HYDROPOWER GENERATION

Background

This dataset was developed for research on anomaly detection in cyber–physical systems (CPSs) such as railways, water treatment plants, and power plants.

In 2017, three laboratory-scale CPS testbeds were initially launched, namely GE's turbine testbed, Emerson's boiler testbed, and FESTO's modular production system (MPS) water treatment testbed. These testbeds were related to relatively simple processes, and were operated independent to each other. In September 2018, a complex process system was built to combine the three testbeds using a HIL simulator, where thermal power generation and pumped-storage hydropower generation were simulated. This ensured that the variables were highly coupled and correlated for a richer dataset. In addition, an open platform communications united architecture (OPC-UA) gateway was installed to facilitate data collection from heterogeneous devices.

The first version of the HAI dataset was made available on GitHub and Kaggle in February 2020. This dataset included ICS operational data from normal and abnormal situations for 38 attacks. Subsequently, a debugged version of HAI v1.0, namely HAI 20.07, was released in July 2020. We newly made HAI v2.0 for the HAICon 2020 competition and a refined version, namely HAI 21.03, was released in March 2021. In 2021, we held an AI-based competition named HAICon 2021. It was an AI-based challenge for industrial control system threat detection. We released the HAI 22.04 version based on the dataset used in the competition.

HAI Testbed

The testbed consisted of a boiler, turbine, water-treatment component, and an HIL simulator. The boiler process involved water-to-water heat transfer based on low pressure and moderate temperature. On the other hand, the turbine process involved closely simulating the behavior of an actual rotating machine using a rotor kit testbed. The boiler and turbine processes were interconnected with the HIL simulator to ensure synchronization with the rotating speed of a steam-power generator. In the water treatment process, water was pumped to the upper reservoir and subsequently released into the lower reservoir according to a pumped-storage hydropower generation model during the HIL simulation.

The three real-world processes, that is, the boiler, turbine, and water treatment processes, were controlled by three different controllers. Emerson Ovation distributed control system (DCS) was used for controlling the water level, flow rate, pressure, temperature, water feed pump, and heater in the boiler process. In the turbine process, GE's Mark VIe DCS was used for speed control and vibration monitoring. A Siemens S7-300 PLC was used in the water treatment process to control the water level and pump. A dSPACE® SCALEXIO system was used for the HIL simulations and interconnected with the real-world processes using a Siemens S7-1500 PLC and ET200 remote IO devices.

TESTBED

Process Architecture

The process flow of the testbed was divided into four primary processes: the boiler process (P1), turbine process (P2), water treatment process (P3), and HIL simulation (P4) (Figure 1). The HIL simulation enhances the correlation between the three real-world processes at the signal level by simulating the thermal power and pumped-storage hydropower generation processes.

The boiler and turbine processes simulated the thermal power plant, while the water treatment process simulated the pumped-storage hydropower plant.

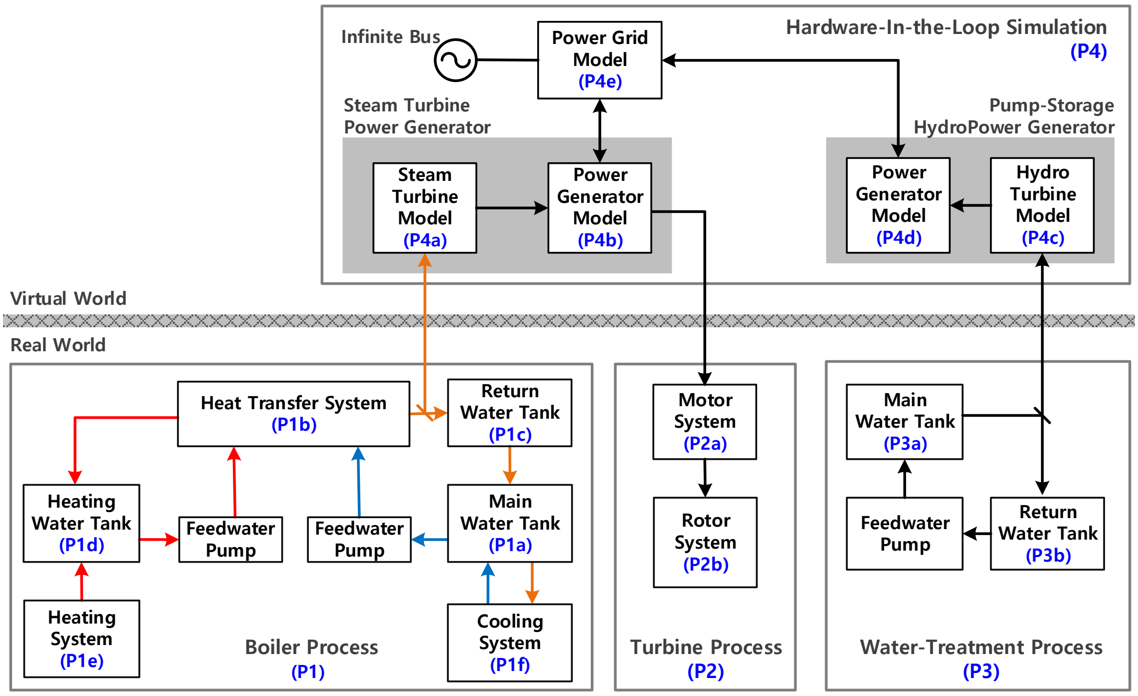

FIGURE 1. PROCESS FLOW DIAGRAM OF THE TESTBED.

P1: BOILER PROCESS

The boiler process involved water-to-water heat transfer at low pressures and moderate temperatures, where the boiler pressure, temperature, and water level are controlled by the boiler process. The opening and closing rates of the main valve are also controlled according to the opening rate of the steam valve of the thermal power plant in the HIL simulator. The pressure and temperature of the main pipe and the water level are transmitted to the HIL simulator in real-time to determine the amount of power generated.

Cool water in the main water tank (P1a) is pumped to the heat-transfer system (P1b) through a feedwater pump, subsequently providing water at a constant temperature and pressure to the return water tank (P1c). The heating system (P1e) transfers thermal energy through the water to the heat transfer system. The water temperature and pressure values are then converted into the current steam temperature and pressure values for the steam-turbine power generator of the HIL simulator (P4a). Water flows from the return water tank (P1c) to the main water tank (P1a) at a constant flow rate, thereby maintaining constant water level in the return water tank. The water circulating to the main tank is not sufficiently cooled; therefore, the cooling system (P1f) additionally removes the thermal energy from the water in the main water tank. The temperature, pressure, level, and flow rate of water in the boiler system were kept constant using eleven sensors, three actuators (two pumps and a heater), and six valves. An operator was able to control five setpoints via the operator workstation (OWS).

P2: TURBINE PROCESS

An actual rotating machine was closely simulated using a GE Rotor Kit (Bently Nevada Asset Condition Monitoring), which consisted of a motor system with a direct-current motor speed control device and a rotor system that allows for coupling and included a rotor shaft, two balance wheels, two journal bearings, and a bearing block. The motor speed was synchronized with the rotating speed of the thermal power generator model in the HIL simulator. The turbine system included a speedometer and four vibration-monitoring proximity probes to maintain a motor speed constant, where the operator can adjust the turbine rotations per minute (RPM) setpoint using a human-machine interface (HMI).

P3: WATER-TREATMENT PROCESS

The water-treatment process involved the pumping and release of water between the upper and lower reservoirs using the hydropower turbine model in the HIL simulation. The water-treatment system included seven sensors, one actuator, and an outflow control valve to control the flow and pressure from the return water tank (P3b) to the main water tank (P3a), as well as the water level in the main water tank. The hydraulic pressure, flow rate, and water level of the upper water tank were transmitted to the HIL simulator in real time to determine the power generation.

P4: HARDWARE-IN-THE-LOOP SIMULATOR

The simulation system consisted of two synchronous generator models (*i.e., steam-turbine power generator and pumped-storage hydropower generator*) and one power grid model, which included the local load demand and was connected to an infinite bus.

An HIL-based simulator was developed to combine the three control systems for the boiler, turbine, and water treatment processes to form a combined power generation system. Specifically, the temperature and pressure of the boiler system were used to determine the pressure and temperature of the steam entering the steam turbine model (STM) (P4.1). The output power of the STM was controlled by an internal steam governor, and the power generator model (P4.2) generated the corresponding electrical power. Further, the hydro turbine model (HTM) (P4.3) and power generator model (P4.4) calculated the generated output power based on the discharge from the water treatment system, where both models were controlled to ensure that the frequency of the microgrid load was 60 Hz (P4.5). The power generated based on the input load was dependent on the opening and closing rates of the valves of the thermal power plant and pumped-storage power plant. Thus, the opening and closing rates of the valves in the control systems for the boiler and water treatment systems were determined.

Testbed Components

The three real-world processes were controlled by three different controllers. Specifically, the boiler process was controlled by Emerson’s Ovation DCS for the water level, flow rate, pressure, temperature, water feed pump, and heater control. The turbine process was controlled by GE’s Mark VIe DCS for speed control and vibration monitoring, and the water treatment process was controlled by a Siemens S7-300 PLC for water level and pump control. In the HAI testbed, the HIL simulations were conducted using a dSPACE® SCALEXIO system interconnected with the real-world processes using a S7-1500 PLC (Siemens) and with an ET200 remote IO devices.

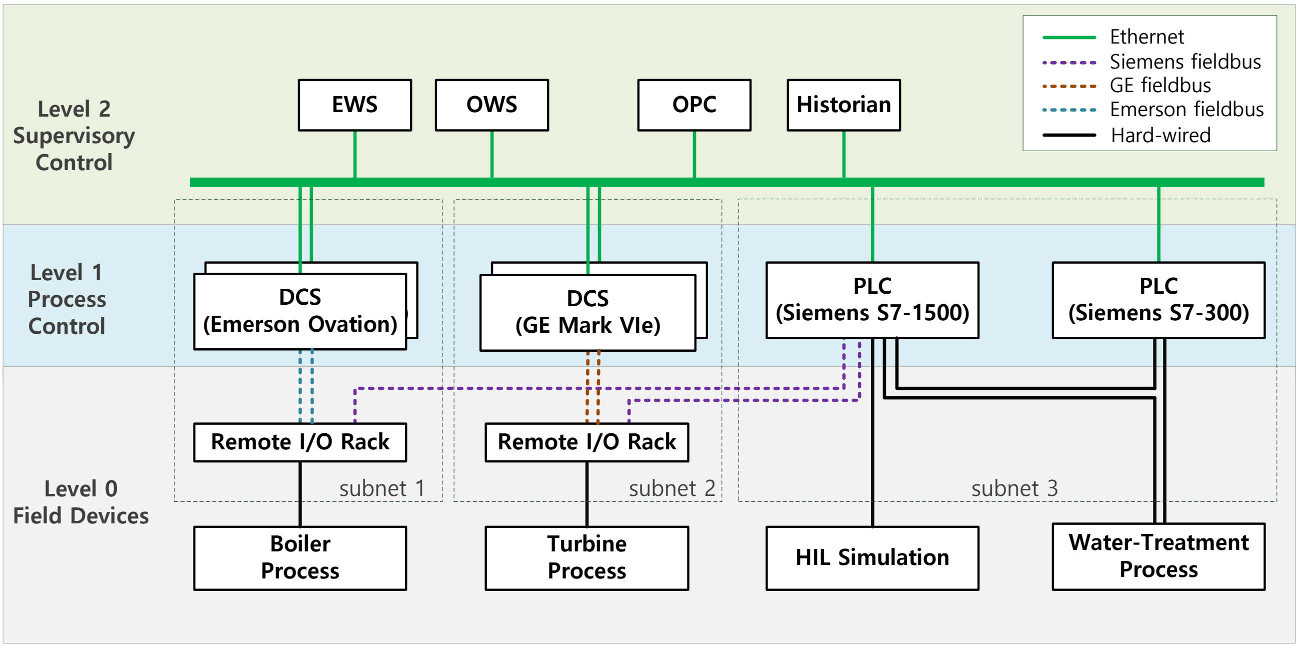

FIGURE 2. TESTBED COMPONENTS AND DATA FLOW.

Process Controllers

BOILER CONTROLLERS

Emerson Ovation DCS consists of four feedback control loops to control the pressure, water level, outflow, temperature, and cooling pump.

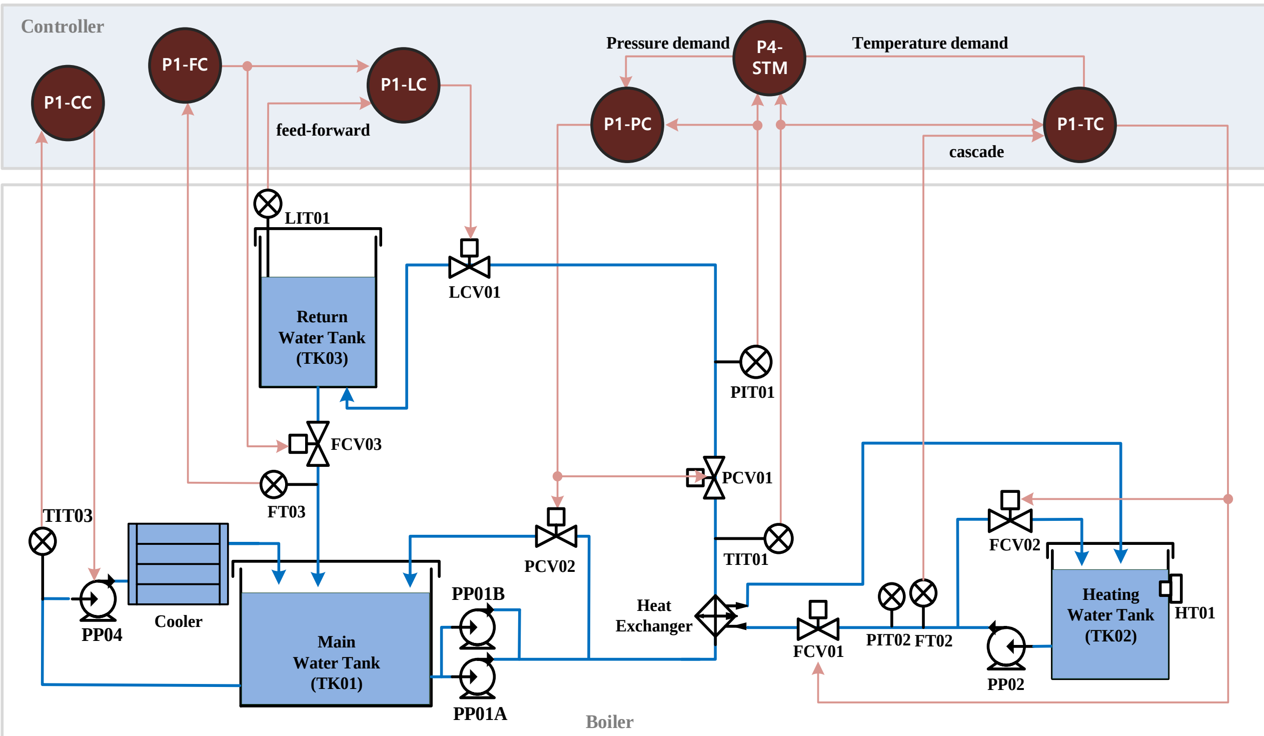

FIGURE 3. BOILER PROCESS CONTROL ARCHITECTURE.

P1-PC: Pressure Controller

P1-PC pressure controller was a feedback controller for two pressure-control valves (PCV01D and PCV02D), and maintained the pressure (PIT01) between the main and return water tanks according to an operator's setpoint command (B2016).

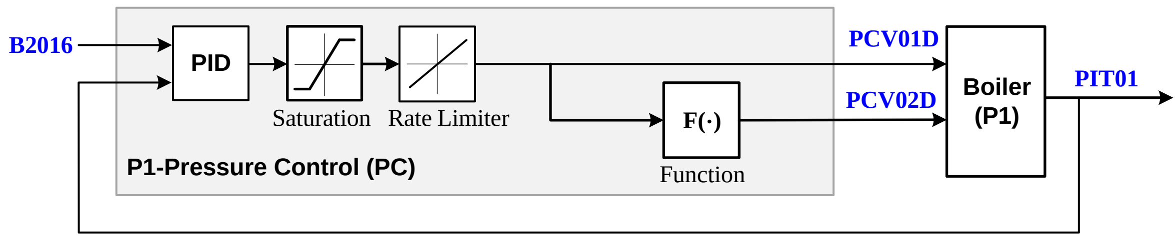

FIGURE 4. PRESSURE CONTROL OF THE BOILER.

P1-LC: Level Controller

P1-LC level controller was a feedback controller for the level-control valve (LCV01D), and maintained the water level (LIT01) of the return water tank according to the operator's setpoint command (B3004). In addition, a feed-forward control was used to rapidly suppress any disturbance in the outflow rate (FCV03D).

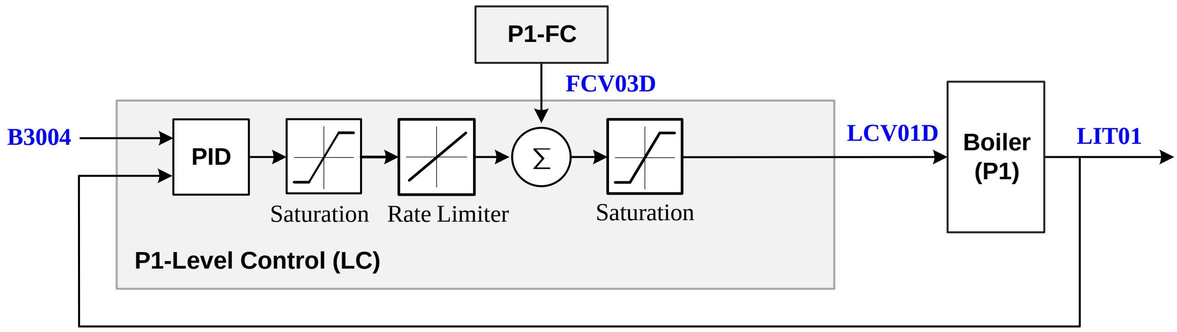

FIGURE 5. LEVEL CONTROL OF THE BOILER.

P1-FC: Flow rate Controller

P1-FC flow rate controller was a feedback controller for the flow-control valve (FCV03D), and maintained the outflow rate (FT03) for the return water tank according to the operator's setpoint command (B3005).

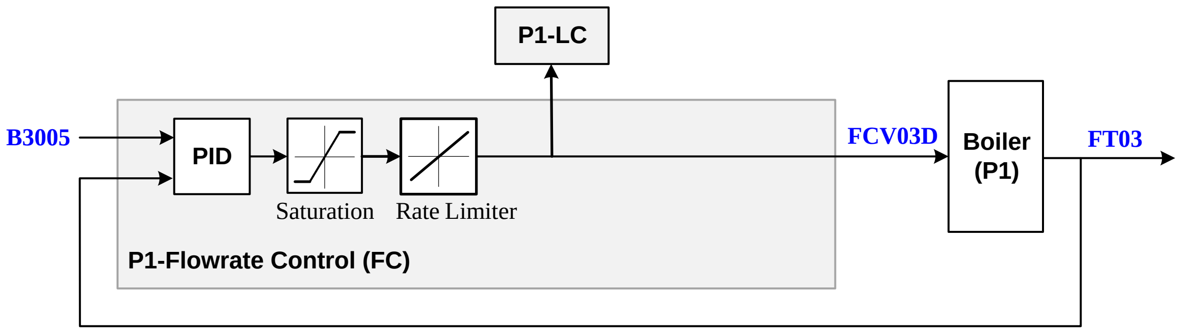

FIGURE 6. FLOW RATE CONTROL OF THE BOILER.

P1-TC: Temperature Controller

P1-TC temperature controller was a feedback controller for two flow-control valves (FCV01D and FCV02D) in the heat transfer system, and maintained the temperature (TIT01) of the main vessel according to the operator's setpoint command (B4022). Cascade control with feedforward compensation to the flow controller (inner loop) based on the water flow allowed for a quicker response to fluctuations in the water flow.

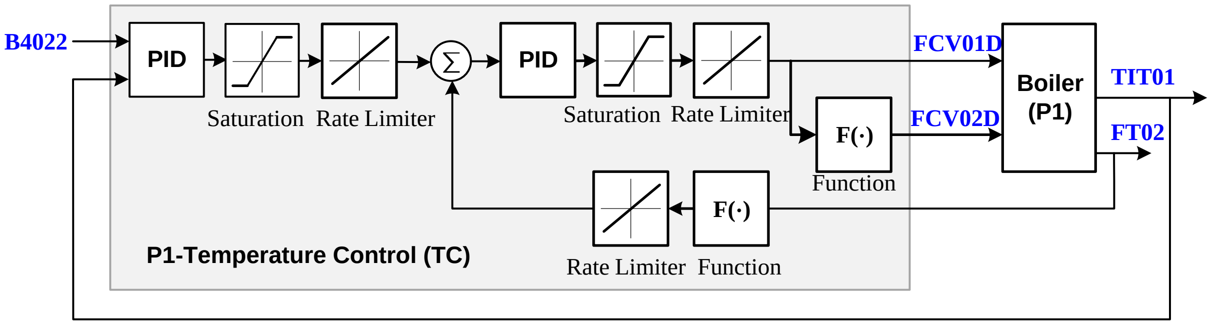

FIGURE 7. TEMPERATURE CONTROL OF THE BOILER.

P1-CC: Cooling Controller

The P1-CC cooling controller drives frequency (PP04) of the cooling water pump. The controller activates the pump operation at the set point (PP04SP) when the water temperature (TIT03) in the main water tank is in the operation range.

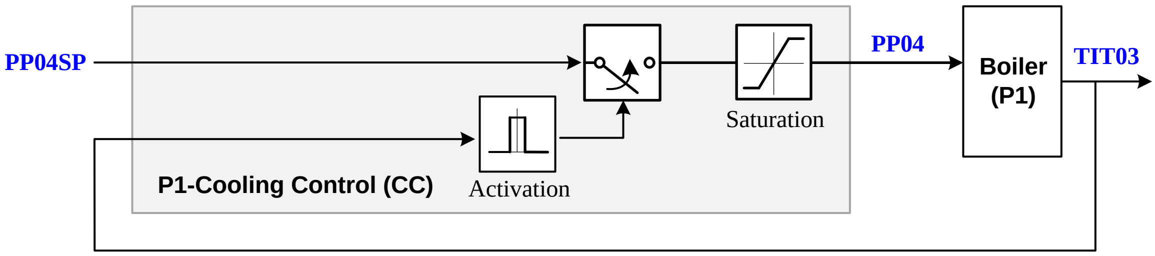

FIGURE 8. COOLING CONTROL OF THE BOILER

TURBINE CONTROLLERS

GE's Mark VIe DCS had one feedback loop that controlled the motor speed. The HIL simulator (P4-STM) generated setpoint trajectories for speed control (P2-SC).

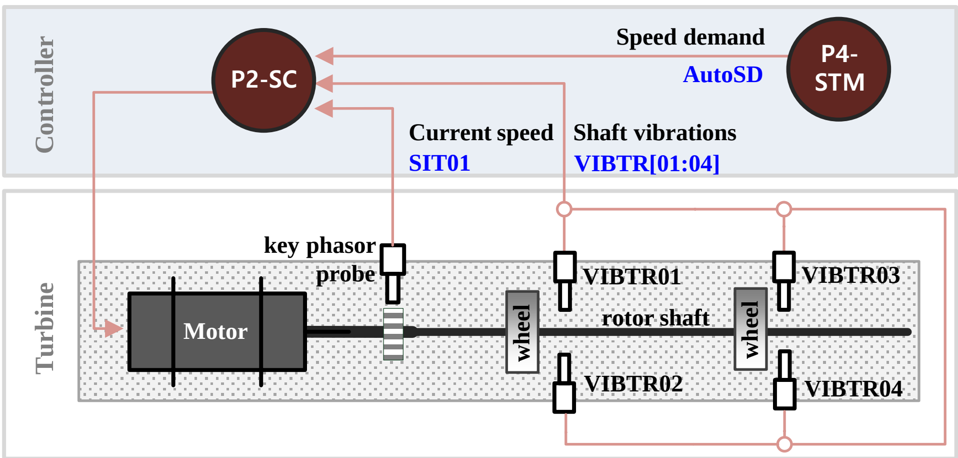

FIGURE 9. TURBINE PROCESS CONTROL ARCHITECTURE.

P2-TRIP: Over-speed and over-vibration trips

The purpose of trip is to prevent an over-speed and over-vibration of a turbine. A turbine runs when the monitored speed (SIT01) is above the RPM trip rate (RTR) or any of four vibration sensors (VIBTR[n]) are above a preset limit (VTR[n]), and then the emergency stop (Emerg) become active. The turbine run mode is activated if the push button to exit the trip mode (TripEx) is successfully triggered.

P2-SC: Speed Control

The P2-SC speed controller increases the motor speed from zero to the minimum controlling speed at a constant rate. Moreover, it facilitates engagement control with a proportional integral derivative (PID) controller to maintain the motor speed value (SIT01) as close as possible to the speed setpoint value (AutoSD).

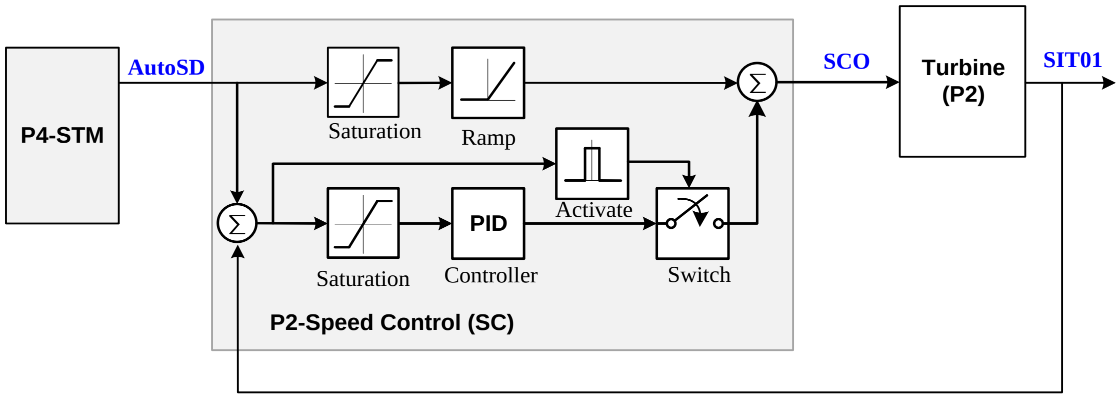

FIGURE 10. SPEED CONTROL OF A TURBINE.

WATER TREATMENT CONTROLLERS

The SIMATIC S7 PCL used for the water treatment control has one feedback loop that controls the water level in the upper reservoir.

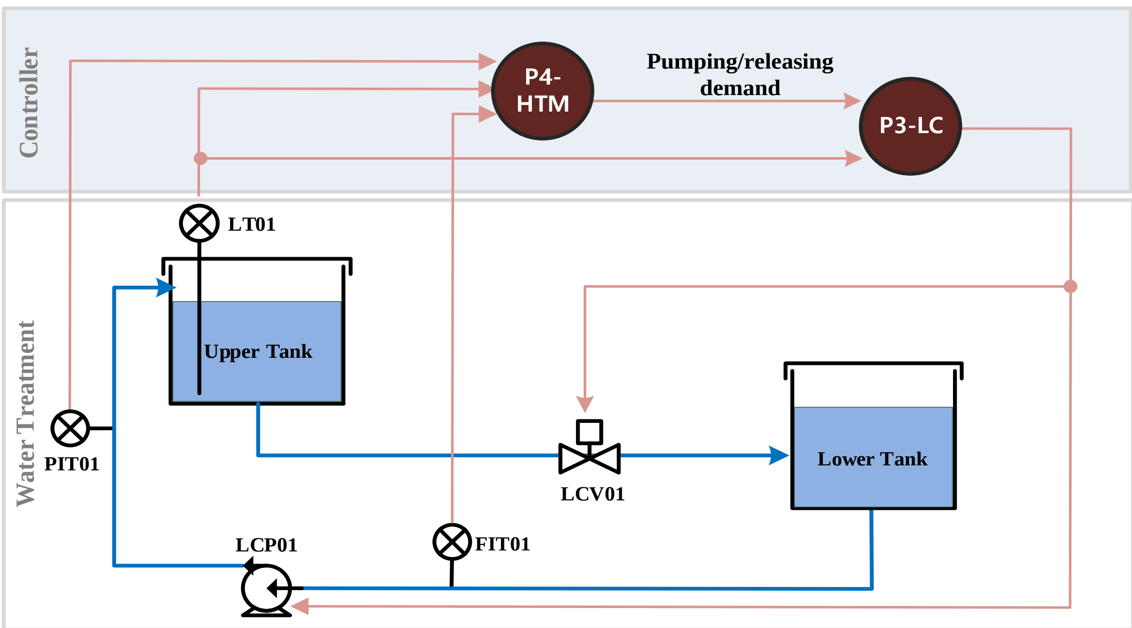

FIGURE 11. WATER TREATMENT PROCESS CONTROL ARCHITECTURE.

P3-LC: Level Control

P3-LC controls the level control valve (LCV01) and level control pump (LCP01) by adjusting the discharge and pumping demands of the HIL simulator.

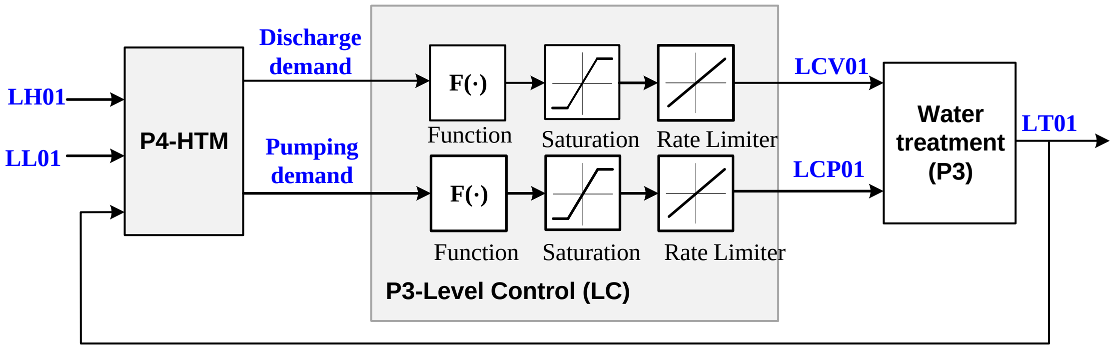

FIGURE 12. WATER LEVEL CONTROL IN A WATER TREATMENT PLANT.

Data Points

All collected data points are tabulated below. Supervisory control and data acquisition (SCADA) systems typically consist of data elements called points (or tags), where each point represents a single variable measured or controlled by the system.

No	Name	Range		Unit	Description	HAI		
		Min	Max			20.07	21.03	22.04
1	P1_B2004	0	10	bar	Heat-exchanger outlet pressure setpoint	√	√	√
2	P1_B2016	0	10	bar	Pressure demand for thermal power output control	√	√	√
3	P1_B3004	0	720	mm	Water level setpoint (return water tank)	√	√	√
4	P1_B3005	0	2,500	l/h	Discharge flowrate setpoint (return water tank)	√	√	√
5	P1_B4002	0	100	℃	Heat-exchanger outlet temperature setpoint	√	√	√
6	P1_B4005	0	100	%	Temperature PID control output	√	√	√
7	P1_B400B	0	2,500	l/h	Water outflow rate setpoint (heating water tank)	√	√	√
8	P1_B4022	0	40	℃	Temperature demand for thermal power output control	√	√	√
9	P1_FCV01D	0	100	%	Position command for the FCV01 valve	√	√	√
10	P1_FCV01Z	0	100	%	Current position of the FCV01 valve	√	√	√
11	P1_FCV02D	0	100	%	Position command for the FCV02 valve	√	√	√
12	P1_FCV02Z	0	100	%	Current position of the FCV02 valve	√	√	√
13	P1_FCV03D	0	100	%	Position command for the FCV03 valve	√	√	√
14	P1_FCV03Z	0	100	%	Current position of the FCV03 valve	√	√	√
15	P1_FT01	0	2,500	mmH2O	Measured flowrate of the return water tank	√	√	√
16	P1_FT01Z	0	3,190	l/h	Water inflow rate converted from P1_FT01	√	√	√

No	Name	Range		Unit	Description	HAI		
		Min	Max			20.07	21.03	22.04
17	P1_FT02	0	2,500	mmH2O	Measured flowrate of heating water tank	√	√	√
18	P1_FT02Z	0	3,190	l/h	Water outflow rate conversion from P1_FT02	√	√	√
19	P1_FT03	0	2,500	mmH2O	Measured flowrate of the return water tank	√	√	√
20	P1_FT03Z	0	3,190	l/h	Water outflow rate converted from P1_FT03	√	√	√
21	P1_LCV01D	0	100	%	Position command for the LCV01 valve	√	√	√
22	P1_LCV01Z	0	100	%	Current position of the LCV01 valve	√	√	√
23	P1_LIT01	0	720	mm	Water level of the return water tank	√	√	√
24	P1_PCV01D	0	100	%	Position command for the PCV01 valve	√	√	√
25	P1_PCV01Z	0	100	%	Current position of the PCV01 valve	√	√	√
26	P1_PCV02D	0	100	%	Position command for the PCV2 valve	√	√	√
27	P1_PCV02Z	0	100	%	Current position of the PCV02 valve	√	√	√
28	P1_PIT01	0	10	bar	Heat-exchanger outlet pressure	√	√	√
29	P1_PIT01_HH	0	10	bar	Highest outlet pressure of the heat-exchanger			√
30	P1_PIT02	0	10	bar	Water supply pressure of the heating water pump	√	√	√
31	P1_PP01AD	0	1	Boolean	Start command of the main water pump PP01A		√	√
32	P1_PP01AR	0	1	Boolean	Running state of the main water pump PP01A		√	√
33	P1_PP01BD	0	1	Boolean	Start command of the main water pump PP01B		√	√
34	P1_PP01BR	0	1	Boolean	Running state of the main water pump PP01B		√	√
35	P1_PP02D	0	1	Boolean	Start command of the heating water pump PP02		√	√
36	P1_PP02R	0	1	Boolean	Running state of the heating water pump PP02		√	√
37	P1_PP04	0	100	%	Control out of the cooler pump			√
38	P1_PP04SP	0	100	℃	Cooler temperature setpoint			√
39	P1_SOL01D	0	1	Boolean	Open command of the main water tank supply valve			√
40	P1_SOL03D	0	1	Boolean	Open command of the main water tank drain valve			√
41	P1_STSP	0	1	Boolean	Start/stop command of the boiler DCS		√	√
42	P1_TIT01	-50	150	℃	Heat-exchanger outlet temperature	√	√	√
43	P1_TIT02	-50	150	℃	Temperature of the heating water tank	√	√	√
44	P1_TIT03	-50	150	℃	Temperature of the main water tank			√

No	Name	Range		Unit	Description	HAI		
		Min	Max			20.07	21.03	22.04
45	P2_24Vdc	0	30	Voltage	DCS 24V Input Voltage	√	√	√
46	P2_ATSW_Lamp	0	1	Boolean	Lamp of the Auto SW			√
47	P2_AutoGo	0	1	Boolean	Auto start button	√ (Auto)	√	√
48	P2_AutoSD	0	3,200	RPM	Auto speed demand	√ (SD01)	√	√
49	P2_Emerg	0	1	Boolean	Emergency button	√ (Emgy)	√	√
50	P2_MASW	0	1	Boolean	Manual(1)/Auto(0) SW			√
51	P2_MASW_Lamp	0	1	Boolean	Lamp of Manual SW			√
52	P2_ManualGO	0	1	Boolean	Manual start button		√	√
53	P2_ManualSD	0	3,200	RPM	Manual speed demand		√	√
54	P2_OnOff	0	1	Boolean	On/off switch of the turbine DCS	√ (On)	√	√
55	P2_RTR	0	2,880	RPM	RPM trip rate		√	√
56	P2_SCO	0	100,000	-	Control output value of the speed controller		√	√
57	P2_SCST	-100	100	RPM	Speed change proportional to frequency change of the STM		√	√
58	P2_SIT01	0	3,200	RPM	Current turbine RPM measured by speed probe	√	√	√
59	P2_TripEx	0	1	Boolean	Trip emergency exit button	√	√	√
60	P2_VIBTR01	-10	10	㎛	Shaft-vibration-related Y-axis displacement near the 1st mass wheel	√ (VYT02)	√	√
61	P2_VIBTR02	-10	10	㎛	Shaft-vibration-related X-axis displacement near the 1st mass wheel	√ (VXT02)	√	√
62	P2_VIBTR03	-10	10	㎛	Shaft-vibration-related Y-axis displacement near the 2nd mass wheel	√ (VYT03)	√	√
63	P2_VIBTR04	-10	10	㎛	Shaft-vibration-related X-axis displacement near the 2nd mass wheel	√ (VXT03)	√	√
64	P2_VT01	11	12	rad/s	Phase lag signal of the key phasor probe	√	√	√
65	P2_VTR01	-10	10	㎛	Preset vibration limit for the sensor P2_VIBTR01		√	√
66	P2_VTR02	-10	10	㎛	Preset vibration limit for the sensor P2_VIBTR02		√	√
67	P2_VTR03	-10	10	㎛	Preset vibration limit for the sensor P2_VIBTR03		√	√
68	P2_VTR04	-10	10	㎛	Preset vibration limit for the sensor P2_VIBTR03		√	√
69	P3_FIT01	0	27,648	-	Flow rate of water flowing into the upper water tank		√	√
70	P3_LCP01D	0	27,648	-	Speed command for the pump LCP01	√	√	√
71	P3_LCV01D	0	27,648	-	Position command for the valve LCV01	√	√	√
72	P3_LH01	0	70	%	High water level set-point	√	√	√

No	Name	Range		Unit	Description	HAI		
		Min	Max			20.07	21.03	22.04
73	P3_LIT01	0	90	%	Water level of the upper water tank	√ (LT01)	√	√
74	P3_LL01	0	70	%	Low water level set-point	√	√	√
75	P3_PIT01	0	27,648	-	Pressure of water flowing into the upper water tank		√	√
76	P4_HT_FD	-0.02	0.02	mHz	Frequency deviation of HTM	√	√	√
77	P4_HT_LD	0	100	MW	Electrical load demand of HTM	√	√	
78	P4_HT_PO	0	100	MW	Output power of HTM	√	√	√
79	P4_HT_PS	0	100	MW	Scheduled power demand of HTM	√	√	√
80	P4_LD	0	500	MW	Total electrical load demand	√	√	√
81	P4_ST_FD	-0.02	0.02	Hz	Frequency deviation of STM	√	√	√
82	P4_ST_GOV	0	27,648	-	Gate opening rate of STM		√	√
83	P4_ST_LD	0	500	MW	Electrical load demand of STM	√	√	√
84	P4_ST_PO	0	500	MW	Output power of STM	√	√	√
85	P4_ST_PS	0	500	MW	Scheduled power demand of STM	√	√	√
86	P4_ST_PT01	0	27,648	-	Digital value of steam pressure of STM	√	√	√
87	P4_ST_TT01	0	27,648	-	Digital value of steam temperature of STM	√	√	√
TOTAL						59	78	86

ATTACK SCENARIOS

All attack scenarios in the viewpoint of a feedback control scheme were configured based on four types of variables, namely the setpoints (SPs), process variables (PVs), control variables (CVs), and control parameters (CPs). An attacker can control all variables by indirectly manipulating any algorithm blocks in the embedded controllers such as the setpoint algorithm, PID controller, signal conditioner and others. Thus, an attacker can ultimately achieve a stealthy attack on the control device.

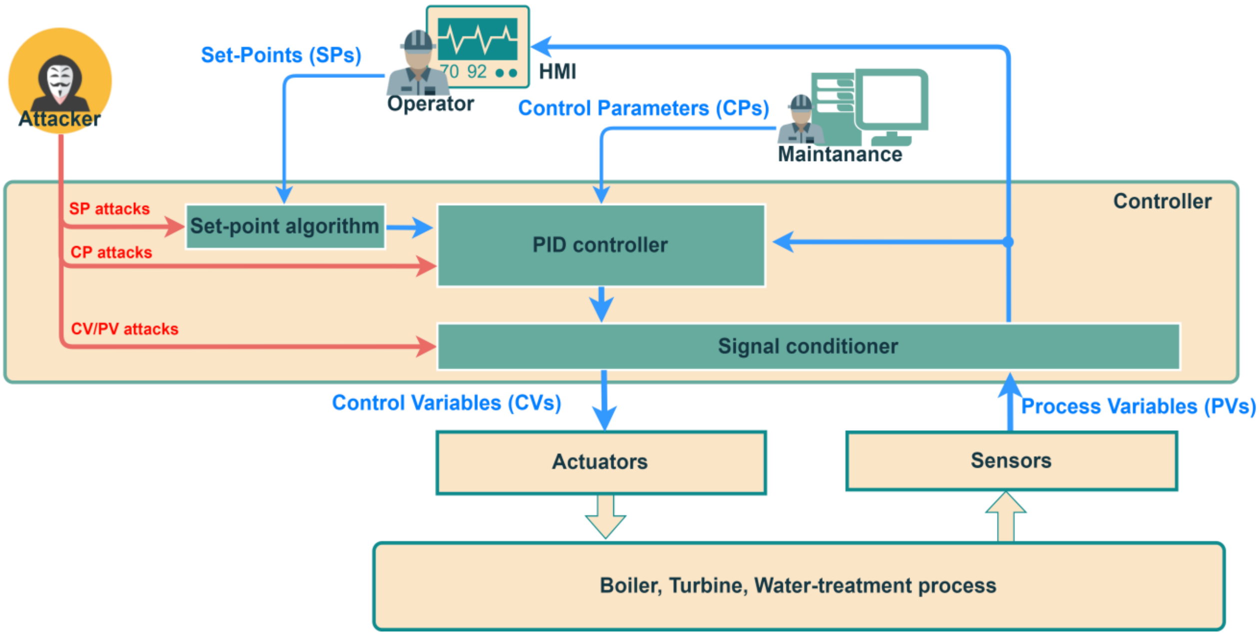

FIGURE 13. ATTACK MODEL BASED ON A PROCESS CONTROL LOOP.

Normal Behaviors

During normal operation, it is assumed that the operator operates the control facility in a routine manner via the HMI, and that the simulator variables associated with power generation in the HIL simulator are changed. The operator monitors the PV values given by the current sensor displayed on the HMI, and adjusts the SPs of the various control devices to operate the system.

HMI operation task scheduler was used to periodically set the SPs and HIL simulator variables to random or predefined values within the normal range to simulate a benign scenario. The normal ranges of SP values in which the entire process was stable were determined by experimentally changing the value of each SP.

The four controllers (P1-PC, P1-LC, P1-FC, and P1-TC) and two simulation models (steam turbine power generator and pump-storage hydropower generator) were automatically operated several times a day. These were initiated with a random delay, and a random value or predefined value within the normal operational range was reached. All SP values were recorded to learn the system features

No	Controller	Set Point	Unit	Normal operational range			
				LowLow	Low	High	HighHigh
1	P1-PC	P1_B2004	bar	0	0.03	0.1	10

2	P1-LC	P1_B3004	mm	0	300	500	720
3	P1-FC	P1_B3005	l/h	0	900	1,100	2,500
4	P1-TC	P1_B4002	℃	0	25	35	100
5	P4-ST	P4_ST_PS	MW	0	0	50	600
6	P4-HT	P4_HT_PS	MW	0	0	50	100

Attack Behaviors

Attack 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.

Since 2019, attack scenarios have been continuously developed, and the attack scenarios have been implemented by considering attack target, attack time, and method for each feedback control loop.

Scenario	Target			Description	HAI		
	Controller	Variable	Point		20.07	21.03	22.04
AP01	P1-PC	SP1	P1_B2016	Decrease or increase SP value of P1-PC. Restore as a form of a trapezoidal profile while hiding SP changes in HMI.	√	√	√
AP02	P1-PC	SP1	P1_B2016	Decrease or increase SP value of P1-PC. Restore as a form of a trapezoidal profile while hiding SP changes in HMI.	√	√	√
		PV1	P1_PIT01	Attempt to maintain previous sensor value.			
AP03	P1-PC	SP1	P1_B2016	Decrease or increase SP value of P1-PC. Restore as a form of a trapezoidal profile while hiding SP changes in HMI.			√
		PV1	P1_PIT01	Attempt to maintain previous sensor value.			
		PV2	P1_FIT01	Attempt to maintain previous sensor value			
AP04	P1-PC	CV1	P1_PCV01D	Decrease or increase CV value of P1-PC. Restore to normal.	√	√	√
AP05	P1-PC	CV1	P1_PCV01D	Decrease or increase CV value of P1-PC. Restore to normal.	√	√	√
		PV1	P1_PIT01	Attempt to maintain previous sensor value.			
AP06	P1-PC	SP1-ST	P1_B2016	Short-term (ST) attack that decrease or increase SP value of P1-PC for a few seconds and restores to normal. Repeat several times while hiding SP changes in HMI.		√	
AP07	P1-PC	CV1-ST	P1_PCV01D	Short-term (ST) attack that decrease or increase CV value of P1-PC for a few seconds and restores to normal. Repeat several times while hiding SP changes in HMI.			√
AP08	P1-FC	SP1	P1_B3005	Decrease or increase SP value of P1-FC. Restore as a form of a trapezoidal profile while hiding SP changes in HMI	√	√	√
AP09	P1-FC	SP1	P1_B3005	Decrease or increase SP value of P1-FC. Restore as a form of a trapezoidal profile while hiding SP changes in HMI	√	√	√
		PV1	P1_FT03	Attempt to maintain previous sensor value.			

Scenario	Target			Description	HAI		
	Controller	Variable	Point		20.07	21.03	22.04
AP10	P1-FC	SP1	P1_B3005	Decrease or increase SP value of P1-FC. Restore as a form of a trapezoidal profile while hiding SP changes in HMI			√
		PV1	P1_FT03	Attempt to maintain previous sensor value.			
		PV2	P1_LIT01	Attempt to maintain previous sensor value.			
AP11	P1-FC	CV1	P1_FCV03D	Decrease or increase CV value of P1-FC. Restore in form of trapezoidal profile.		√	√
AP12	P1-FC	CV1	P1_FCV03D	Decrease or increase CV value of P1-FC. Restore to normal.		√	√
		PV1	P1_FT03	Attempt to maintain previous sensor value.			
AP13	P1-FC	CV1-ST	P1_FCV03D	Short-term (ST) attack that decrease or increase CV value of P1-FC for a few seconds and restores to normal. Repeat several times while hiding SP changes in HMI.		√	√
AP14	P1-LC	SP1	P1_B3004	Decrease or increase SP value of P1-LC. Restore as a form of a trapezoidal profile while hiding SP changes in HMI.	√	√	√
AP15	P1-LC	SP1	P1_B3004	Decrease or increase SP value of P1-LC. Restore as a form of a trapezoidal profile while hiding SP changes in HMI.	√	√	√
		PV1	P1_LIT01	Attempt to repeat previous sensor value.			
AP16	P1-LC	CV1	P1_LCV01D	Decrease or increase CV value of P1-LC. Restore to normal.	√	√	√
AP17	P1-LC	CV1	P1_LCV01D	Decrease or increase CV value of P1-LC. Restore to normal.	√	√	√
		PV1	P1_LIT01	Attempt to repeat previous sensor value.			
AP18	P1-LC	CV1-ST	P1_LCV01D	Short-term (ST) attack that decrease or increase CV value of P1-LC for a few seconds and restores to normal. Repeat several times while hiding SP changes in HMI.		√	√
AP19	P1-TC	CV1	P1_FCV01D	Decrease or increase CV value of P1-TC. Restore to normal.			√
AP20	P1-TC	CV1	P1_FCV01D	Decrease or increase CV value of P1-TC. Restore to normal.			√
		PV1	P1_TIT01	Attempt to repeat previous sensor value.			
AP21	P1-TC	CV1-ST	P1_FCV01D	Short-term (ST) attack that decrease or increase CV value of P1-TC for a few seconds and restores to normal. Repeat several times while hiding SP changes in HMI.			√
AP22	P1-TC	SP1-LT	P1_B4002	Long-term (LT) attack that decrease or increase SP value of P1-TC continuously for more than 10 minutes and restores to normal.			√
AP23	P1-CC	CV1	P1_PP04	Decrease or increase CV value of P1-CC. Restore to normal.			√
AP24	P1-CC	CV1-ST	P1_PP04	Short-term (ST) attack that decrease or increase CV value of P1-CC for a few seconds and restores to normal. Repeat several times while hiding SP changes in HMI.			√

Scenario	Target			Description	HAI		
	Controller	Variable	Point		20.07	21.03	22.04
AP25	P1-CC	SP1-LT	P1_PP04_SP	Long-term (LT) attack that decrease or increase SP value of P1-CC continuously for more than 10 minutes and restores to normal.			√
AP26	P2-SC	SP1	P2_AutoSD (P2_SD01)	Decrease or increase SP value of P2-SC. Restore as a form of a trapezoidal profile while hiding SP changes in HMI.	√	√	√
AP27	P2-SC	SP1	P2_AutoSD (P2_SD01)	Decrease or increase SP value of P2-SC. Restore as a form of a trapezoidal profile while hiding SP changes in HMI.	√	√	√
		PV1	P2_SIT01	Attempt to maintain previous sensor value.			
AP28	P2-SC	SP2	P2_ManualSD	Decrease or increase SP value of P2-SC. Restore as a form of a trapezoidal profile while hiding SP changes in HMI.			√
AP29	P2-SC	CV1	P2_SCO	Decrease or increase CV value of P2-SC. Restore to normal.		√	√
AP30	P2-SC	CV1	P2_SCO	Decrease or increase CV value of P2-SC. Restore to normal.		√	√
		PV1	P2_SIT01	Attempt to maintain previous sensor value.			
AP31	P2-SC	SP1-ST	P2_AutoSD	Short-term (ST) attack that decrease or increase CV value of P2-SC for a few seconds and restores to normal. Repeat several times while hiding SP changes in HMI.		√	√
AP32	P2-TC	SP1	P2_VTR01	Decrease or increase SP value of P2-TC. Restore as a form of a trapezoidal profile while hiding SP changes in HMI.		√	
AP33	P2-TC	SP2	P2_VTR02	Decrease or increase SP value of P2-SC. Restore as a form of a trapezoidal profile while hiding SP changes in HMI.		√	√
AP34	P2-TC	SP3	P2_RTR	Decrease or increase SP value of P2-SC. Restore as a form of a trapezoidal profile while hiding SP changes in HMI.		√	√
AP35	P3-LC	CV1	P3_LCP01D	Attempt to repeat previous sensor value.	√	√	√
AP36	P3-LC	CV1	P3_LCP01D	Decrease or increase CV value of P3-LC. Restore to normal.			√
		PV1	P3_LIT01	Attempt to maintain previous sensor value.			
AP37	P3-LC	CV2	P3_LCV01D	Decrease or increase CV value of P3-LC. Restore to normal.	√	√	√
AP38	P3-LC	CV2	P3_LCV01D	Decrease or increase CV value of P3-LC. Restore to normal.			√
		PV1	P3_LIT01	Attempt to maintain previous sensor value.			
AP39	P3-LC	CV2-LT	P3_LCV01D	Long-term (LT) attack that decrease or increase CV value of P3-LC continuously for more than 10 minutes and restores to normal.			√
TOTAL					**14**	**25**	**37**

DATASETS

Since 2020, three versions of the dataset have been released, and herein, these datasets are described in detail starting with latest version. It is noteworthy that the version numbering follows a date-based scheme, where the version number indicates the released date.

HAI 22.04

HAI 22.04 includes six CSV files as training datasets and four CSV files as testing datasets. The time-series data in each CSV file satisfies time continuity and includes 89 columns. The first column represents the observed time in the “yyyy-MM-dd hh:mm:ss” format, while the next 87 columns provide the recorded SCADA data points. The last four columns provide data labels for the presence or absence of an attack. Out of these columns, the attack column is applicable to all processes and the other three columns are applicable to the corresponding control processes.

NORMAL OPERATION

We recently used a hidden Markov model (HMM) to model the normal operation of SCADA. The HMM probabilistically determines the sequence and the delivery time of set point commands from a set of seven set points. Three HMMs are constructed to generate normal operations of three process controllers of the HAI testbed. The internal states and transition probability were constructed by considering the general process of each process control. The set-points are finally output probabilistically as possible observations. The probabilistic parameters of all the HMMs were given below. The change value of each observation was randomly determined within its normal range.

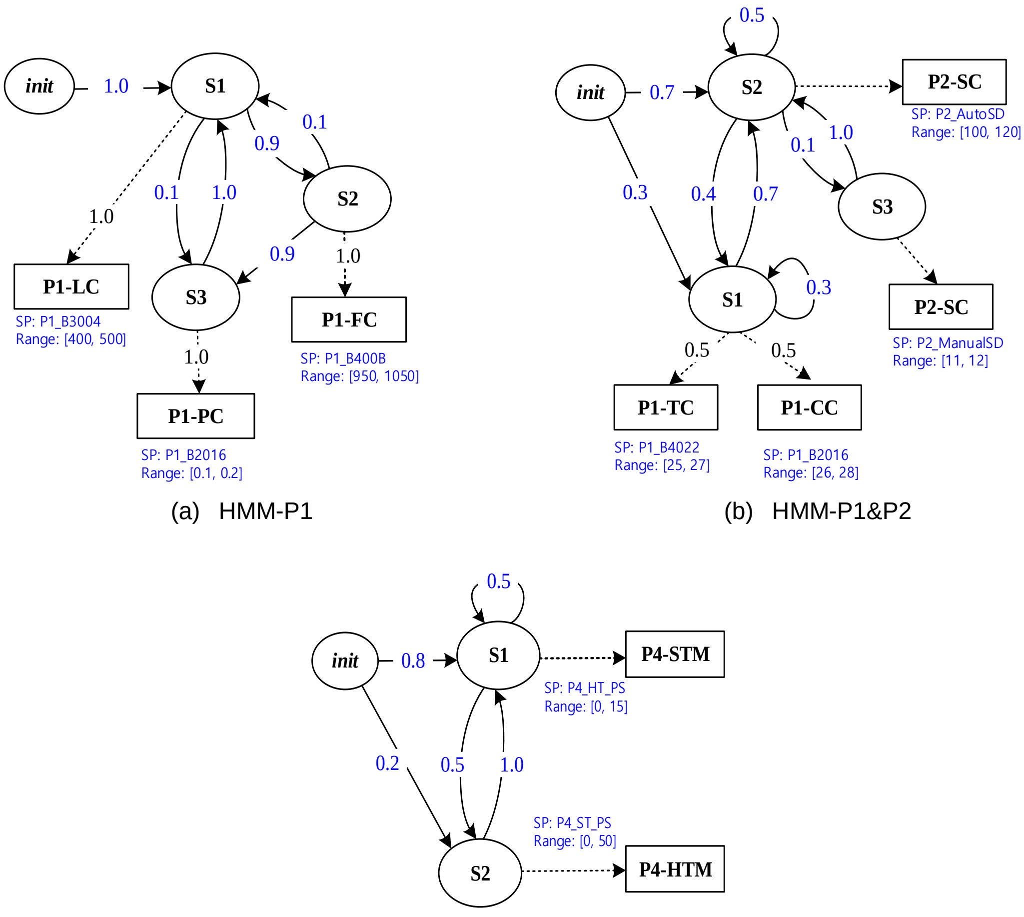

FIGURE 14. HMM-BASED GENERATIVE MODELS FOR NORMAL OPERATION.

ATTACK OPERATION

The 58 attacks were conducted, including 32 attack primitives and 26 combinations of attacks designed to simultaneously perform two attack primitives. The attack scenarios are given below.

No	ID	Attack			Start Time		Duration (sec)
		Scenario	Target Controller	Target Point(s)			
1	A101	AP04	P1-PC-CO1	P1_PCV01D	Jul. 10, 2021	5:41	190
2	A102	AP18	P1-LC-CO1-ST	P1_LCV01D		7:19	54
3	A103	AP11	P1-FC-CO1PV1	P1_FCV03D, P1_FT03		11:25	126
4	A104	AP37	P3-LC-CO2	P3_LCV01D		15:39	54
5	A105	AP14	P1-LC-SP1	P1_B3004		16:42	296
6	A106	AP13	P1-CC-CO1	P1_PP04		19:21	91
7	A107	AP19	P1-TC-CO1	P1_FCV01D		22:35	67
8	A201	AP01	P1-PC-SP1	P1_B2016	Jul. 13, 2021	16:38	257
9	A202	AP13	P1-FC-CO1-ST	P1_FCV03D		17:21	65
10	A203	AP31	P2-SC-SP1-ST	P2_AutoSD		18:13	45
11	A204	AP04	P1-PC-CO1	P1_PCV01D		20:28	248
		AP29	P2-SC-CO1	P2_SCO			
12	A205	AP37	P3-LC-CO2	P3_LCV01D		21:10	55
13	A206	AP02	P1-PC-SP1PV1	P1_B2016, P1_PIT01		21:58	176
		AP27	P2-SC-SP1PV1	P2_AutoSD, P2_SIT01			
14	A207	AP16	P1-LC-CO1	P1_LCV01D		23:40	284
15	A208	AP30	P2-SC-CO1PV1	P2_SCO, P2_SIT01	Jul. 14, 2021	1:15	152
16	A209	AP03	P1-PC-SP1PV1PV2	P1_B2016, P1_PIT01, P1_FIT01		1:40	162
17	A210	AP26	P2-SC-SP1	P2_AutoSD		3:23	97
18	A211	AP05	P1-PC- CO1PV1	P1_PCV01D, P1_PIT01		7:21	151
19	A212	AP35	P3-LC-CO1	P3_LCP01D		8:11	55
20	A213	AP24	P1-CC-CO1-ST	P1_PP04		10:35	80
21	A214	AP39	P3-LC-CO2-LT	P3_LCV01D		11:23	613
22	A215	AP09	P1-FC-SP1PV1	P1_B3005, P1_FT03		12:17	168
23	A216	AP01	P1-PC-SP1	P1_B2016		13:52	158
		AP08	P1-FC-SP1	P1_B3005			

No	ID	Attack			Start Time		Duration (sec)
		Scenario	Target Controller	Target Point(s)			
24	A217	AP10	P1-FC-CO1	P1_FCV03D		14:31	98
25	A301	AP16	P3-LC-CO2	P2_LCV01D		18:21	348
		AP10	P1-FC-CO1	P1_FCV03D			
26	A302	AP15	P1-LC-SP1PV1	P1_LCV01D		20:16	358
27	A303	AP17	P1-LC-CO1PV1	P1_B3004. P1_LIT01		23:22	143
		AP37	P3-LC-CO2	P3_LCV01D			
28	A304	AP38	P3-LC-CO2PV1	P1_LCV01D. P1_LIT01		1:41	91
29	A305	AP18	P1-LC-CO1-ST	P3_LCV01D		2:09	94
30	A306	AP04	P1-PC-CO1	P1_LCV01D		3:37	353
		AP15	P1-LC-SP1PV1	P1_B3004. P1_LIT01			
31	A307	AP20	P1-TC-CO1PV1	P1_FCV01D. P1_TIT01		5:35	151
32	A308	AP05	P1-PC-CO1PV1	P1_PCV01D, P1_PIT01	Jul. 15, 2021	6:53	173
		AP23	P1-CC-CO1	P1_PP04			
33	A309	AP08	P1-FC-SP1	P1_B3005		7:42	96
		AP19	P1-TC-CO1	P1_FCV01D			
34	A310	AP35	P3-LC-CO1	P3_LCP01D		9:52	2024
		AP37	P3-LC-CO2	P3_LCV01D			
35	A401	AP28	P2-SC-SP2	P2_ManualSD		12:42	38
36	A402	AP21	P1-TC-CO1-ST	P1_FCV01D		13:20	88
37	A403	AP34	P2-TC-SP3	P2_RTR		13:57	96
38	A404	AP26	P2-SC-SP1	P2_AutoSD		15:08	97
		AP37	P3-LC-CO2	P3_LCV01D			
39	A405	AP22	P1-TC-SP1-LT	P1_B4002		16:07	505
40	A406	AP09	P1-FC-SP1PV1	P1_B3005, P1_FT03		17:22	186
		AP19	P1-TC-CO1	P1_FCV01D			
41	A407	AP13	P1-FC-CO1-ST	P1_FCV03D		19:45	122
		AP17	P1-LC-CO1PV1	P1_LCV01D, P1_LIT01			
42	A408	AP05	P1-PC-CO1PV1	P1_PCV01D, P1_PIT01		20:29	673
		AP17	P1-LC-CO1PV1	P1_LCV01D, P1_LIT01			

No	ID	Attack			Start Time		Duration (sec)
		Scenario	Target Controller	Target Point(s)			
43	A409	AP18	P1-LC-CO1-ST8	P1_LCV01D		22:41	63
		AP21	P1-TC-CO1-ST9	P1_FCV01D			
44	A410	AP11	P1-FC-CO1PV1	P1_FCV03D, P1_FT03		01:07	179
		AP27	P2-SC-SP1PV1	P2_AutoSD, P2_SIT01			
45	A411	AP23	P1-CC-CO1	P1_PP04		03:35	99
		AP34	P2-TC-SP3	P2_RTR			
	A412	AP20	P1-TC-CO1PV1	P1_FCV01D, P1_TIT01		04:02	156
		AP01	P1-PC-SP1	P1_B2016			
47	A413	AP16	P1-LC-CO1	P1_LCV01D		04:59	153
		AP27	P2-SC-SP1PV1	P2_AutoSD, P2_SIT01			
48	A414	AP33	P2-TC-SP2	P2_VTR02		07:20	77
		AP36	P3-LC-CO1PV1	P3_LCP01D, P3_LIT01			
49	A415	AP3	P2-TC-SP2	P2_VTR02		09:17	77
50	A416	AP12	P1-FC-CO1PV1PV2	P1_FCV03D, P1_FT03, P1_LIT01		10:39	134
51	A417	AP25	P1-CC-SP1-LT	P1_PP04_SP.	Jul. 16, 2021	11:22	544
52	A418	AP01	P1-PC-SP1	P1_B2016		13:23	342
		AP14	P1-LC-SP1	P1_B3004			
53	A419	AP01	P1-PC-SP1	P1_B2016		14:59	163
		AP35	P3-LC-CO1	P3_LCP01D			
54	A420	AP07	P1-PC-CO1-ST	P1_PCV01D		15:57	89
55	A421	AP30	P2-SC-CO1PV1	P2_SCO, P2_SIT01		17:34	152
		AP23	P1-CC-CO1	P1_PP04			
56	A422	AP02	P1-PC-SP1PV1	P1_B2016, P1_PIT01		20:08	165
		AP26	P2-SC-SP1	P2_AutoSD			
57	A423	AP08	P1-FC-SP1	P1_B3005		22:17	115
		AP29	P2-SC-CO1	P2_SCO			
58	A424	AP10	P1-FC-CO1	P1_FCV03D		23:05	86
		AP23	P1-CC-CO1	P1_PP04			

HAI 21.03

HAI 21.03 includes three CSV files as training datasets and five CSV files as testing datasets. The time-series data in each CSV file satisfies time continuity, and includes 84 columns. The first column represents the observed time as “yyyy-MM-dd hh:mm:ss,” while 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 process and the other three columns were for the corresponding control processes.

NORMAL OPERATION

An HMI operation task scheduler periodically sets the SPs and HIL simulator variables to predefined values within the normal range to simulate a benign scenario. The benign scenarios are given below.

No	Set points						Start Time
	P1_B2004 (Pressure SP)	P1_B3004 (Level SP)	P1_B3005 (Flowrate SP)	P1_B4002 (Temperature SP)	P4_ST_PS (Scheduled Power)	P4_HT_PS (Scheduled Power)	
1	0.1 (±0.002)	440 (±9)	1,100 (±22)	32 (±0)	0 (±0)	0 (±0)	03:00 (±10)
2	0.03 (±0.001)	400 (±8)	1,100 (±22)	32 (±0)	0 (±0)	0 (±0)	04:30 (±10)
3	0.1 (±0.002)	400 (±8)	1,100 (±22)	32 (±1)	0 (±0)	0 (±0)	06:00 (±10)
4	0.1 (±0.002)	400 (±8)	900 (±18)	32 (±0)	0 (±0)	0 (±0)	08:30 (±10)
5	0.1 (±0.002)	380 (±8)	1,100 (±22)	32 (±0)	0 (±0)	0 (±0)	10:00 (±10)
6	0.06 (±0.001)	420 (±8)	1,000 (±20)	32 (±0)	0 (±0)	0 (±0)	12:00 (±0)
7	0.1 (±0.002)	400 (±40)	1,100 (±22)	32 (±0)	0 (±0)	0 (±0)	14:30 (±10)
8	0.1 (±0.002)	400 (±8)	1,000 (±60)	33 (±1)	0 (±0)	0 (±0)	17:00 (±10)
9	0.1 (±0.002)	400 (±8)	1,100 (±22)	32 (±1)	0 (±0)	0 (±0)	19:30 (±10)
10	0.1 (±0.002)	400 (±8)	1,100 (±22)	32 (±1)	50 (±0)	10 (±0)	22:00 (±10)

ATTACK OPERATION

The 50 attacks were conducted, including 25 attack primitives and 25 combinations of attacks designed to simultaneously perform two attack primitives. The attack scenarios are given below.

No	ID	Attack			Start Time		Duration (sec)
		Scenario	Target Controller	Target Point(s)			
1	A101	AP01	P1-PC-SP1	P1_B2016	Jul. 7, 2020	15:35	192
2	A102	AP06	P1-FC-SP1	P1_B3005		17:28	98
3	A103	AP13	P1-LC-CO1	P1_LCV01D		18:59	190
4	A104	AP18	P2-SC-CO1	P2_SCO		20:21	60
5	A105	AP16	P2-SC-SP1	P2_AutoSD		21:03	89
6	A201	AP22	P2-TC-SP2	P2_VTR02	Jul. 9, 2020	15:47	83
7	A202	AP02	P1-PC-SP1PV1	P1_B2016, P1_PIT01		17:38	422
8	A203	AP15	P1-LC-CO1-ST7	P1_LCV01D		18:59	17

No	ID	Attack			Start Time		Duration (sec)
		Scenario	Target Controller	Target Point(s)			
9	A204	AP07	P1-FC-SP1PV1	P1_B3005, P1_FT03		20:10	259
10	A205	AP05	P1-PC-SP1-ST10	P1_B2016		21:15	123
11	A206	AP09	P1-FC-CO1PV1	P1_FCV03D, P1_FT03		23:02	256
12	A207	AP21	P2-TC-SP1	P2_VTR01	Jul. 10, 2020	01:08	68
13	A208	AP12	P1-LC-SP1PV1	P1_B3004, P1_LIT01		01:33	261
14	A209	AP11	P1-LC-SP1	P1_B3004		03:03	159
15	A210	AP04	P1-PC-CO1PV1	P1_PCV01D, P1_PIT01		05:29	421
16	A211	AP20	P2-SC-SP1-ST5	P2_AutoSD		07:51	45
17	A212	AP17	P2-SC-SP1PV1	P2_AutoSD, P2_SIT01		09:13	152
18	A213	AP14	P1-LC-CO1PV1	P1_LCV01D, P1_LIT01		10:49	254
19	A214	AP03	P1-PC-CO1	P1_PCV01D		12:51	152
20	A215	AP19	P2-SC-CO1PV1	P2_SCO, P2_SIT01		15:11	151
21	A216	AP10	P1-FC-CO1-ST10	P1_FCV03D		15:40	65
22	A217	AP23	P2-TC-SP3	P2_RTR		16:22	184
23	A218	AP08	P1-FC-CO1	P1_FCV03D		18:21	99
24	A219	AP24	P3-LC-CO1	P3_LCP01D		21:25	119
25	A220	AP25	P3-LC-CO2	P2_LCV01D		22:56	119
26	A301	AP15	P1-LC-CO1-ST	P1_LCV01D	Jul. 13, 2020	13:51	132
		AP06	P1-FC-SP1	P1_B3005			
27	A302	AP02	P1-PC-SP1PV1	P1_B2016, P1_PIT01		15:21	421
		AP06	P1-FC-SP1	P1_B3005			
28	A303	AP03	P1-PC-CO1	P1_PCV01D		18:11	189
		AP13	P1-LC-CO1	P1_LCV01D			
29	A304	AP16	P2-SC-SP1	P2_AutoSD		20:53	106
		AP21	P2-TC-SP1	P2_VTR01			
30	A305	AP18	P2-SC-CO1	P2_SCO		21:23	84
		AP22	P2-TC-SP2	P2_VTR02			
31	A306	AP01	P1-PC-SP1	P1_B2016		23:55	238
		AP16	P2-SC-SP1	P2_AutoSD			
32	A307	AP08	P1-FC-CO1	P1_FCV03D	Jul. 14, 2020	01:51	110
		AP21	P2-TC-SP1	P2_VTR01			
33	A308	AP14	P1-LC-CO1PV1	P1_LCV01D, P1_LIT01		03:53	255
		AP20	P2-SC-SP1-ST	P2_AutoSD			
34	A401	AP03	P1-PC-CO1	P1_PCV01D	Jul. 28,	12:43	254

No	ID	Attack			Start Time	Duration (sec)
		Scenario	Target Controller	Target Point(s)		
		AP12	P1-LC-SP1PV1	P1_B3004, P1_LIT01	2020	
35	A402	AP07	P1-FC-SP1PV1	P1_B3005, P1_FT03		13:45	262
		AP25	P3-LC-CO2	P2_LCV01D			
36	A403	AP12	P1-LC-SP1PV1	P1_B3004, P1_LIT01		15:57	263
		AP25	P3-LC-CO2	P2_LCV01D			
37	A404	AP19	P2-SC-CO1PV1	P2_SCO, P2_SIT01		17:45	258
		AP14	P1-LC-CO1PV1	P1_LCV01D, P1_LIT01			
38	A405	AP20	P2-SC-SP1-ST	P2_AutoSD		20:47	120
		AP25	P3-LC-CO2	P2_LCV01D			
39	A501	AP03	P1-PC-CO1	P1_PCV01D	Jul. 30, 2020	11:16	172
		AP22	P2-TC-SP2	P2_VTR02			
40	A502	AP09	P1-FC-CO1PV1	P1_FCV03D, P1_FT03		13:30	258
		AP18	P2-SC-CO1	P2_SCO			
41	A503	AP12	P1-LC-SP1PV1	P1_B3004, P1_LIT01		16:05	256
		AP18	P2-SC-CO1	P2_SCO			
42	A504	AP08	P1-FC-CO1	P1_FCV03D		17:45	120
		AP25	P3-LC-CO2	P2_LCV01D			
43	A505	AP11	P1-LC-SP1	P1_B3004		18:38	203
		AP20	P2-SC-SP1-ST	P2_AutoSD			
44	A506	AP19	P2-SC-CO1PV1	P2_SCO, P2_SIT01		20:42	153
		AP25	P3-LC-CO2	P2_LCV01D			
45	A507	AP20	P2-SC-SP1-ST	P2_AutoSD		23:13	79
		AP21	P2-TC-SP1	P2_VTR01			
46	A508	AP10	P1-FC-CO1-ST	P1_FCV03D	Jul. 31, 2020	01:15	51
		AP15	P1-LC-CO1-ST	P1_LCV01D			
47	A509	AP01	P1-PC-SP1	P1_B2016		02:01	241
		AP03	P1-PC-CO1	P1_PCV01D			
48	A510	AP11	P1-LC-SP1	P1_B3004		09:54	262
		AP14	P1-LC-CO1PV1	P1_LCV01D, P1_LIT01			
49	A511	AP23	P2-TC-SP3	P2_RTR		10:40	120
		AP25	P3-LC-CO2	P2_LCV01D			
50	A512	AP06	P1-FC-SP1	P1_B3005		11:21	262
		AP09	P1-FC-CO1PV1	P1_FCV03D, P1_FT03			

HAI 20.07

HAI 20.07 includes two CSV files as training datasets and two CSV files as testing datasets. The time-series data in each CSV file satisfies time continuity and includes 63 columns. The first column represents the observed time in the "yyyy-MM-dd hh:mm:ss" format, and the remaining 59 columns provide the recorded SCADA data points. The last four columns provide data labels for whether an attack occurred or not. Out of these columns, the attack column is applicable to all processes and the other three columns are applicable to the corresponding control processes.

NORMAL OPERATION

The normal operations of the first training dataset (train1.csv) are given below, where all the SP change commands were delivered at the start of each day.

No	Setpoint						Start Time
	P1_B2004 (Pressure SP)	P1_B3004 (Level SP)	P1_B3005 (Flowrate SP)	P1_B4002 (Temperature SP)	P4_ST_PS (Scheduled Power)	P4_HT_PS (Scheduled Power)	
1	0.1 (±0.002)	460 (±20)	1,100 (±22)	32 (±0)	0 (±0)	0 (±0)	7:00 (±0)
2	0.03 (±0.002)	400 (±8)	1,100 (±22)	32 (±0)	0 (±0)	0 (±0)	9:00 (±0)
3	0.1 (±0.002)	400 (±8)	1,100 (±22)	31 (±1)	0 (±0)	0 (±0)	11:00 (±0)
4	0.1 (±0.002)	400 (±8)	1,000 (±100)	32 (±0)	0 (±0)	0 (±0)	13:00 (±0)
5	0.1 (±0.002)	400 (±8)	1,100 (±22)	32 (±0)	50 (±5)	0 (±0)	15:00 (±0)

The normal operations of the second training dataset (train2.csv) are given below.

No	Setpoint						Start Time
	P1_B2004 (Pressure SP)	P1_B3004 (Level SP)	P1_B3005 (Flowrate SP)	P1_B4002 (Temperature SP)	P4_ST_PS (Scheduled Power)	P4_HT_PS (Scheduled Power)	
1	0.03 (±0.002)	400 (±8)	1,100 (±22)	32 (±0)	0 (±0)	0 (±0)	00:00 (±0)
2	0.1 (±0.002)	450 (±20)	1,100 (±22)	32 (±0)	0 (±0)	0 (±0)	10:00 (±0)
3	0.1 (±0.002)	400 (±8)	1,100 (±22)	32 (±1)	0 (±0)	0 (±0)	14:00 (±0)
4	0.1 (±0.002)	400 (±8)	1,000 (±100)	32 (±0)	0 (±0)	0 (±0)	16:00 (±0)
5	0.1 (±0.002)	400 (±8)	1,100 (±22)	32 (±0)	50 (±5)	0 (±0)	22:00 (±0)

ATTACK OPERATION

A total of 38 attacks were conducted, including 14 attack primitives and 14 combinations of attacks designed to simultaneously perform two attack primitives.

No	ID	Attack			Start Time		Duration (sec)
		Scenario	Target Controller	Target Point(s)			
1	A101	AP12	P1-LC-SP1PV1	P1_B3004, P1_LIT01	Oct. 29, 2019	13:40	370
2	A102	AP13	P1-LC-CV1	P1_LCV01D		14:35	312

No	ID	Attack			Start Time		Duration (sec)
		Scenario	Target Controller	Target Point(s)			
3	A103	AP14	P1-LC-CV1PV1	P1_LCV01D, P1_LIT01		15:45	868
4	A104	AP06	P1-FC-SP1	P1_B3005		16:30	262
5	A105	AP11	P1-LC-SP1	P1_B3004	Oct. 30, 2019	08:50	371
6	A106	AP01	P1-PC-SP1	P1_B2016		09:40	334
7	A107	AP02	P1-PC-SP1PV1	P1_B2016, P1_PIT01		10:35	504
8	A108	AP03	P1-PC-CV1	P1_PCV01D		11:37	268
9	A109	AP04	P1-PC-CV1PV1	P1_PCV01D, P1_PIT01		12:30	518
10	A110	AP17	P2-SC-SP1PV1	P2_SD01, P2_SIT01		14:30	370
11	A111	AP26	P3-LC-SP1CV1	P3_LH01, P3_LCP01		15:35	180
12	A112	AP27	P3-LC-SP2CV2	P3_LL01, P3_LCV01		16:33	154
13	A113	AP16	P2-SC-SP1	P2_SD01	Oct. 31, 2019	08:42	348
14	A114	AP17	P2-SC-SP1PV1	P2_SD01, P2_SIT01		10:30	518
		AP02	P1-PC-SP1PV1	P1_B2016, P1_PIT01			
15	A115	AP16	P2-SC-SP1	P2_SD01		11:33	346
		AP03	P1-PC-CV1	P1_PCV01D			
16	A116	AP17	P2-SC-SP1PV1	P2_SD01, P2_SIT01		13:25	368
17	A117	AP17	P2-SC-SP1PV1	P2_SD01, P2_SIT01		14:30	396
		AP14	P1-LC-CV1PV1	P1_LCV01D, P1_LIT01			
18	A118	AP16	P2-SC-SP1	P2_SD01		15:41	348
		AP06	P1-FC-SP1	P1_B3005			
19	A119	AP26	P3-LC-SP1CV1	P3_LH01, P3_LCP01		16:29	398
		AP01	P1-PC-SP1	P1_B2016			
20	A201	AP26	P3-LC-SP1CV1	P3_LH01, P3_LCP01	Nov. 1, 2019	09:29	560
		AP12	P1-LC-SP1PV1	P1_B3004, P1_LIT01			
21	A202	AP26	P3-LC-SP1CV1	P3_LH01, P3_LCP01		10:41	310
		AP13	P1-LC-CV1	P1_LCV01D			
22	A203	AP26	P3-LC-SP1CV1	P3_LH01, P3_LCP01		11:23	180
23	A204	AP11	P1-LC-SP1	P1_B3004		12:31	506

No	ID	Attack			Start Time		Duration (sec)
		Scenario	Target Controller	Target Point(s)			
		AP07	P1-FC-SP1PV1	P1_B3005, P1_FT03			
24	A205	AP03	P1-PC-CV1	P1_PCV01D		13:41	580
		AP07	P1-FC-SP1PV1	P1_B3005, P1_FT03			
25	A206	AP01	P1-PC-SP1	P1_B2016		14:23	310
26	A207	AP02	P1-PC-SP1PV1	P1_B2016, P1_PIT01		15:31	520
		AP06	P1-FC-SP1	P1_B3005			
27	A208	AP07	P1-FC-SP1PV1	P1_B3005, P1_FT03		16:18	560
28	A209	AP27	P3-LC-SP2CV2	P3_LL01, P3_LCV01		17:20	520
		AP02	P1-PC-SP1PV1	P1_B2016, P1_PIT01			
29	A210	AP01	P1-PC-SP1	P1_B2016	Nov. 4, 2019	15:31	410
		AP06	P1-FC-SP1	P1_B3005			
30	A211	AP24	P3-LC-SP2CV2	P3_SP02, P3_LCV01		17:20	520
		AP01	P1-PC-SP1	P1_B2016			
31	A212	AP24	P3-LC-SP2CV2	P3_SP02, P3_LCV01	Nov. 5, 2019	09:30	380
		AP13	P1-LC-CV1	P1_LCV01D			
32	A213	AP24	P3-LC-SP2CV2	P3_SP02, P3_LCV01		10:20	290
		AP06	P1-FC-SP1	P1_B3005			
33	A214	AP16	P2-SC-SP1	P2_SD01		11:23	340
34	A215	AP16	P2-SC-SP1	P2_SD01		12:30	340
		AP27	P3-LC-SP2CV2	P3_LL01, P3_LCV01			
35	A216	AP16	P2-SC-SP1	P2_SD01		14:45	2,880
		AP11	P1-LC-SP1	P1_B3004			
36	A217	AP11	P1-LC-SP1	P1_B3004		16:20	330
		AP01	P1-PC-SP1	P1_B2016			
37	A218	AP13	P1-LC-CV1	P1_LCV01D		17:23	310
38	A219	AP13	P1-LC-CV1	P1_LCV01D	Nov. 6, 2019	08:58	310
		AP03	P1-PC-CV1	P1_PCV01D			

CITING THE DATASET

Please cite the sources below if you are referencing any of the HAI datasets, performance matric, and competitions. Please do not hesitate to share your results with us.

Datasets

[HAI 22.04] Hyeok-Ki Shin, Woomyo Lee, Jeong-Han Yun and Byung-Gil Min, "ICS security dataset", 2022. *GitHub*, Available at: https://github.com/icsdataset.

[HAI 21.03] Hyeok-Ki Shin, Woomyo Lee, Jeong-Han Yun and Byung-Gil Min, "Two ICS Security Datasets and Anomaly Detection Contest on the HIL-based Augmented ICS Testbed," *In Cyber Security Experimentation and Test (CSET `21)*, Association for Computing Machinery, pp.36-40, 2021.

[HAI 20.07] Hyeok-Ki Shin, Woomyo Lee, Jeong-Han Yun and Hyoungchun Kim, "HAI 1.0: HIL-based Augmented ICS Security Dataset," *13th USENIX Workshop on Cyber Security Experimentation and Test (CSET 20)*, Santa Clara, CA, 2020.

- https://github.com/icsdataset/hai
- https://kaggle.com/icsdataset/hai-security-dataset

Performance Metric

[eTaPR] Won-Seok Hwang, Jeong-Han Yun, Jonguk Kim, and Byung Gil Min, "Do You Know Existing Accuracy Metrics Overate Time-Series Anomaly Detection?", SAC 2022: Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing, 2022.

- https://github.com/saurf4ng/eTapR

Competitions/Baseline

[HAICon] We held an AI contest, namely HAICon, to revitalized research, discover ideas, and improve HAI dataset more. You can find the winner's codes and baseline codes on the official website below.

- HAICon 2021: https://dacon.io/en/competitions/official/235757/codeshare
- HAICon 2020: https://dacon.io/en/competitions/official/235624/codeshare

ABBREVIATIONS

C

CV	CONTROL VARIABLE
CC	COOLING CONTROLLER

D

DCS	DISTRIBUTED CONTROL SYSTEM

F

FC	FLOW CONTROLLER
FCV	FLOW CONTROL VALVE
FIT	FLOW INDICATOR TRANSMITTER
FT	FLOW TRANSMITTER

H

HH	HIGH HIGH
HIL	HARDWARE-IN-THE-LOOP
HMI	HUMAN MACHINE INTERFACE

L

LC	LEVEL CONTROLLER
LCV	LEVEL CONTROL VALVE
LIT	LEVEL INDICATOR TRANSMITTER
LL	LOW LOW
LLH	LIQUID LEVEL [HIGH]
LLL	LIQUID LEVEL [LOW]
LLN	LIQUID LEVEL [NORMAL]
LSH	LEVEL SWITCH [HIGH]
LSHL	EVEL SWITCH [HIGH/LOW]
LSL	LEVEL SWITCH [LOW]
LT	EVEL TRANSMITTER

P

PC	PRESSURE CONTROLLER
PCL	PROCESS CONTROL LOOP
PCV	PRESSURE CONTROL VALVE
PIT	PRESSURE INDICATOR TRANSMITTER
PLC	PROGRAMMABLE LOGIC CONTROLLER
PV	PROCESS VARIABLE

S

SC	SPEED CONTROLLER

SI	**SPEED INDICATOR**
SIT	**SPEED-INDICATOR TRANSMITTER**
SP	**SETPOINT**
SS	**STEAM SUPPLY**

T

TCV	**TEMPERATURE CONTROL VALVE**
TIT	**TEMPERATURE-INDICATOR TRANSMITTER**
TT	**TEMPERATURE TRANSMITTER**

V

VT	**VIBRATION TRANSMITTER**