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

Document Version: 2.0
Release Date: March 2021

RELEASE HISTORY

HAI is a security dataset including both the normal and abnormal behaviors for ICS anomaly detection research. The normal dataset was collected continuously for several days, and 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 thus far. Each dataset consists of several CSV files, and each file satisfies time continuity. The quantitative summary of each version are as follows:

Release Version	*Data Points*	*Normal Dataset*			*Abnormal Dataset*			
		Files	*Interval (hour)*	*Size (MB)*	*Files*	*Attack Count*	*Interval (hour)*	*Size (MB)*
HAI 21.03	***78 points/sec***	*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*
HAI 20.07	***59 points/sec***	*train1.csv*	*86*	*127*	*test1.csv*	*28*	*81*	*119*
		train2.csv	*91*	*98*	*test2.csv*	*10*	*42*	*62*

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

DOCUMENT CHANGE LOG

Version	*Release Date*	*Comments*	*Page*
v2.0	***Feb. 17, 2020***	***Major revision for HAI 21.03***	
		+ Add a brief description of Turbine's trip control	*06*
		+ Add 20 more points in data point table	*08 – 10*
		+ Add 11 more attack scenarios	*11 – 12*
		+ Remove description related to multiple attacks	*12*
		+ Add details of HAI 21.03	*14 – 16*
		+ Update changes to HAI 20.07	*17 – 19*
v1.1	***Jul. 22, 2020***	***Minor revision for HAI 20.07***	
		+ Change the version numbering scheme	*All*
		+ Update range and description of data points	*08 – 10*
		+ Add time duration in attack timetable	*17 – 19*
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, 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 independently of each other. In September 2018, a complex process system was built to combine the three systems using a hardware-in-the-loop (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 HAI dataset was made available on GitHub and Kaggle in February 2020. This dataset included ICS operational data from both normal and anomalous situations for 38 attacks. Subsequently, a debugged version of HAI v1.0, namely HAI 20.07, was released for the HAICon 2020 competition in August 2020. HAI 21.03 was released in 2021, and is based on a more tightly coupled HIL simulator to produce clearer attack effects with additional attacks. This provided more quantitative information and covers a variety of operational situations and better insights into the dynamic changes of the physical system.

HAI TESTBED

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

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

TESTBED OVERVIEW

PROCESS ARCHITECTURE

The process flow of the testbed was divided into four primary processes, namely 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 thermal power generation and pumped-storage hydropower generation scenarios.

The boiler and turbine processes were used to simulate the thermal power plant, while the water treatment process was used to simulate the pumped-storage hydropower plant.

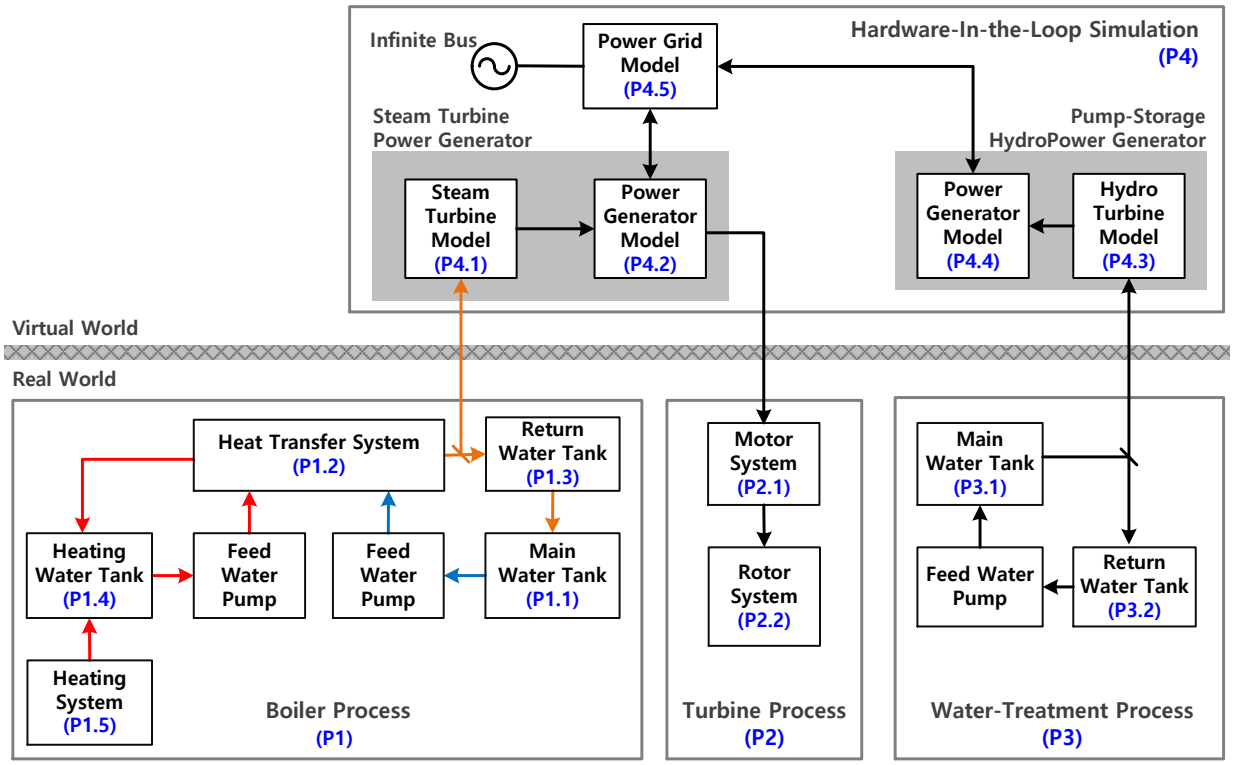

FIGURE 1. PROCESS FLOW DIAGRAM.

P1: Boiler Process

The boiler process involved water-to-water heat transfer at low pressure and moderate temperature, where the boiler process controls the boiler pressure, temperature, and water level. The opening and closing rates of the main valve were also controlled based on 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 were transmitted to the HIL simulator in real time to determine the amount of power generated.

Water was pumped from the main water tank to the heat-transfer system (P1.2), which subsequently provided water at a constant temperature and pressure to the return water system (P1.3). The water temperature and pressure were then converted into the current steam temperature and pressure values for the steam-turbine power generator of the HIL simulator (P4.1). Water was returned to the

main water tank maintain a constant water level in the return water tank (P1.3). The water temperature, water pressure, water level, and flow rate of the boiler system were maintained 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 (P3.2) to the main water tank (P3.1), as well as the water level in the main water tank (P3.1). 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 Simulation

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.

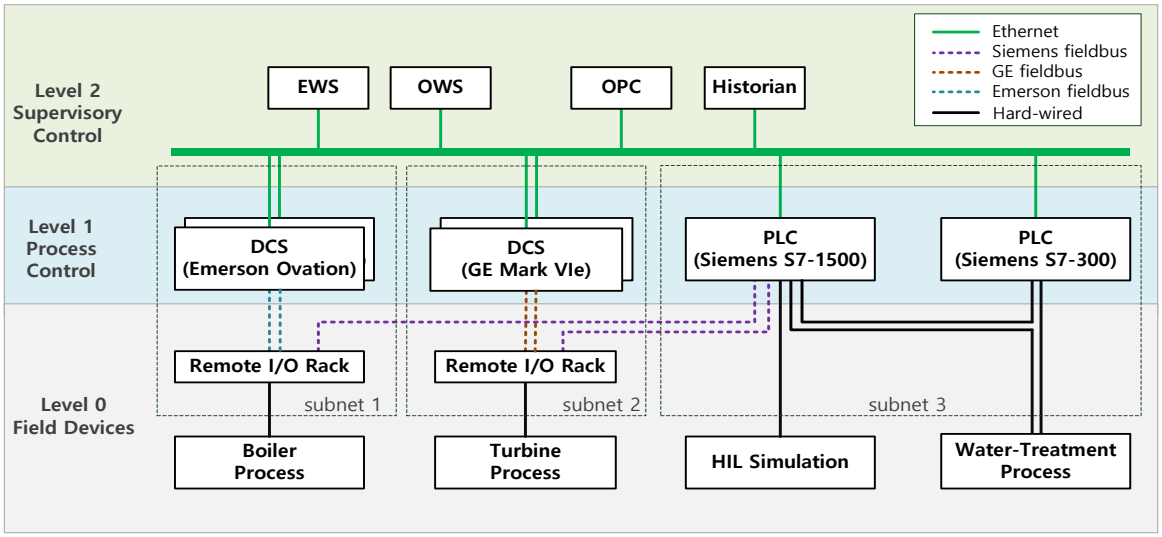

FIGURE 2. TESTBED COMPONENTS AND DATA FLOW.

PROCESS CONTROLLERS

P1: Boiler Controllers

Emerson's Ovation DCS had four feedback loops to control the pressure, water level, outflow, and temperature.

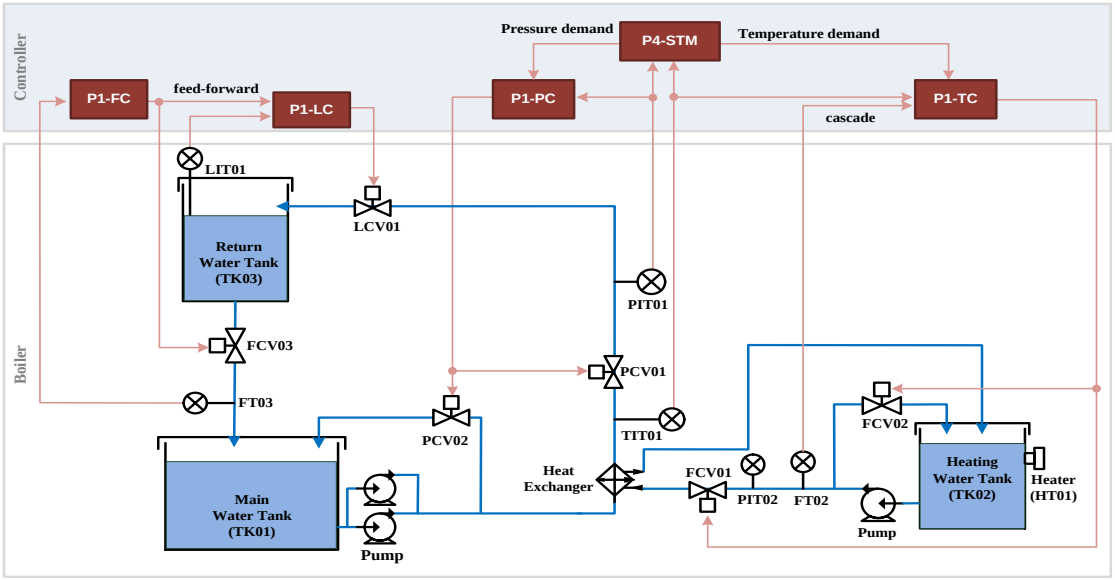

FIGURE 3. ARCHITECTURE OF THE BOILER PROCESS.

P1-PC: Pressure Control

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).

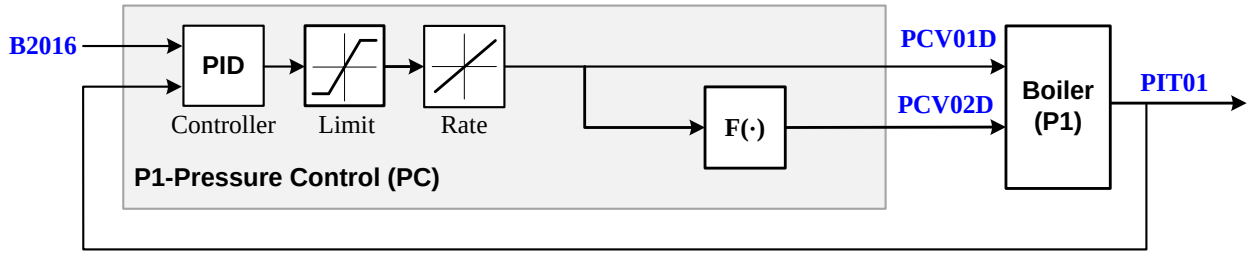

FIGURE 4. PRESSURE CONTROL OF THE BOILER.

P1-LC: Level Control

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).

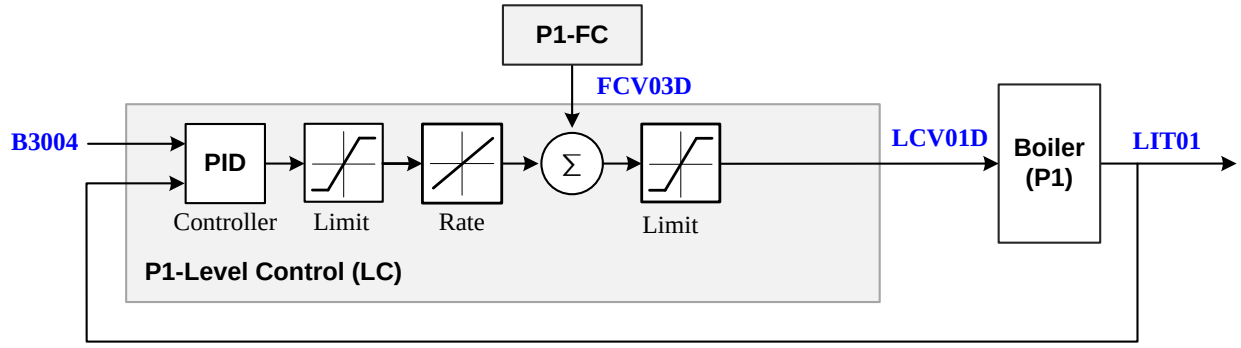

FIGURE 5. LEVEL CONTROL OF THE BOILER.

P1-FC: Flow Rate Control

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).

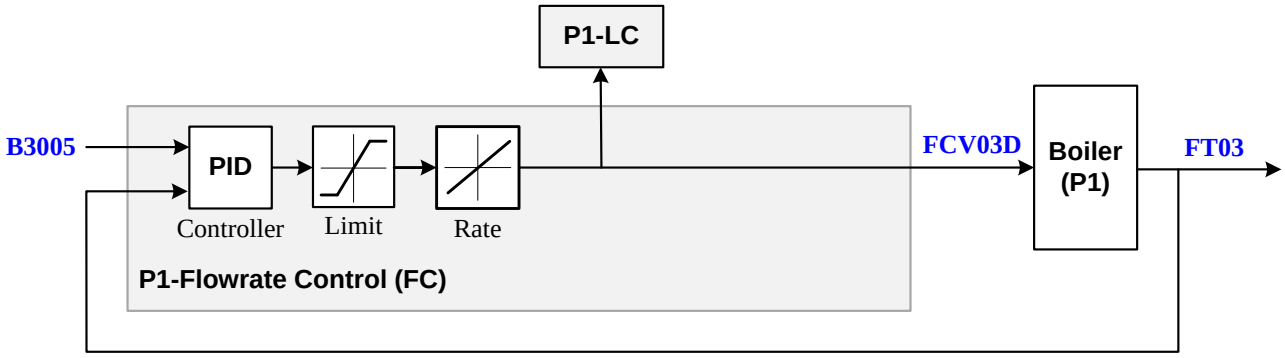

FIGURE 6. FLOW RATE CONTROL OF THE BOILER.

P1-TC: Temperature Control

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.

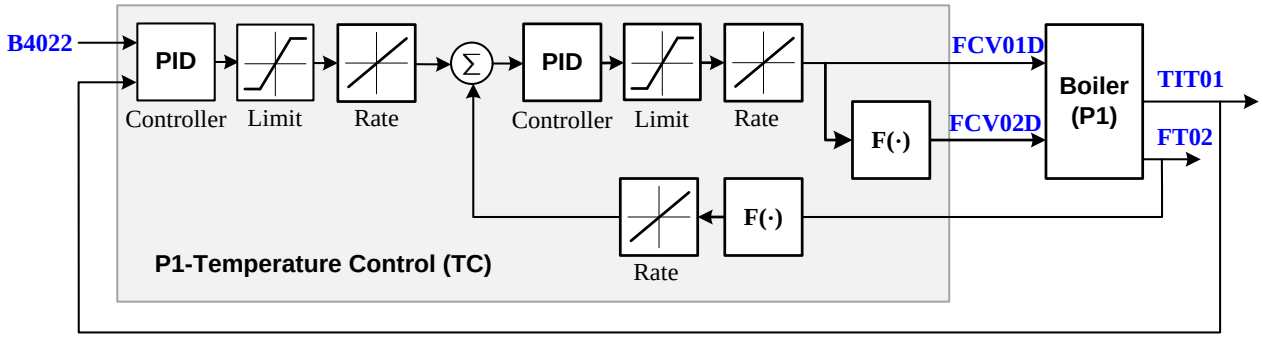

FIGURE 7. TEMPERATURE CONTROL OF THE BOILER.

P2: 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).

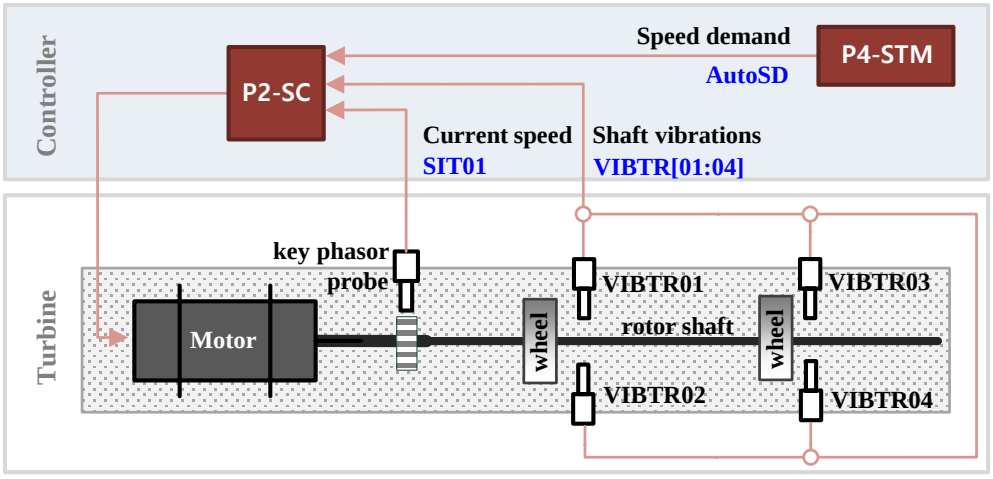

FIGURE 8. PROCESS ARCHITECTURE OF THE TURBINE.

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

The purpose of TRIP is to prevent an overspeed and over-vibration of turbine. Turbine is running 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 emergency stop (Emerg) is active. Turbine run mode is activated if TripEx is successfully triggered.

P2-SC: Speed Control

P2-SC speed controller increased the motor speed from zero to the minimum controlling speed at a constant rate, and facilitated engagement control with a proportional integral derivative (PID) controller to maintain a motor speed (SIT01) as close as possible to the speed setpoint (AutoSD).

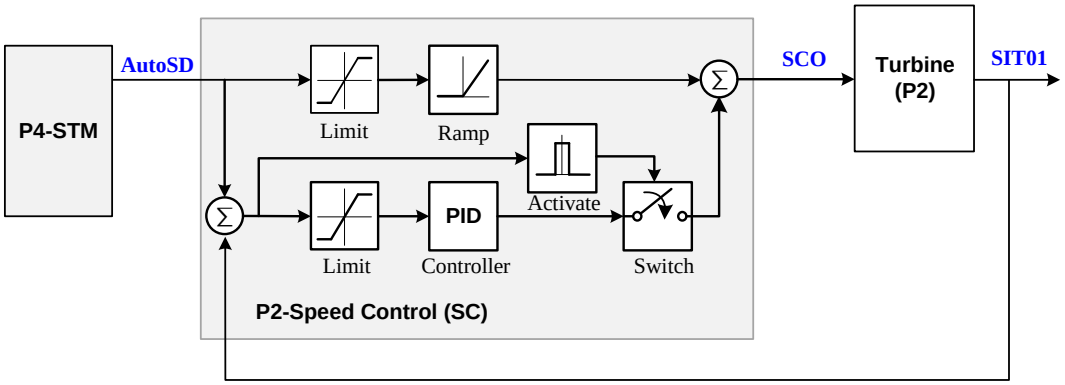

FIGURE 9. SPEED CONTROL OF TURBINE.

P3: Water-Treatment Controllers

The SIMATIC S7 PCL used for water treatment control had one feedback loop that controlled the water level of the upper reservoir.

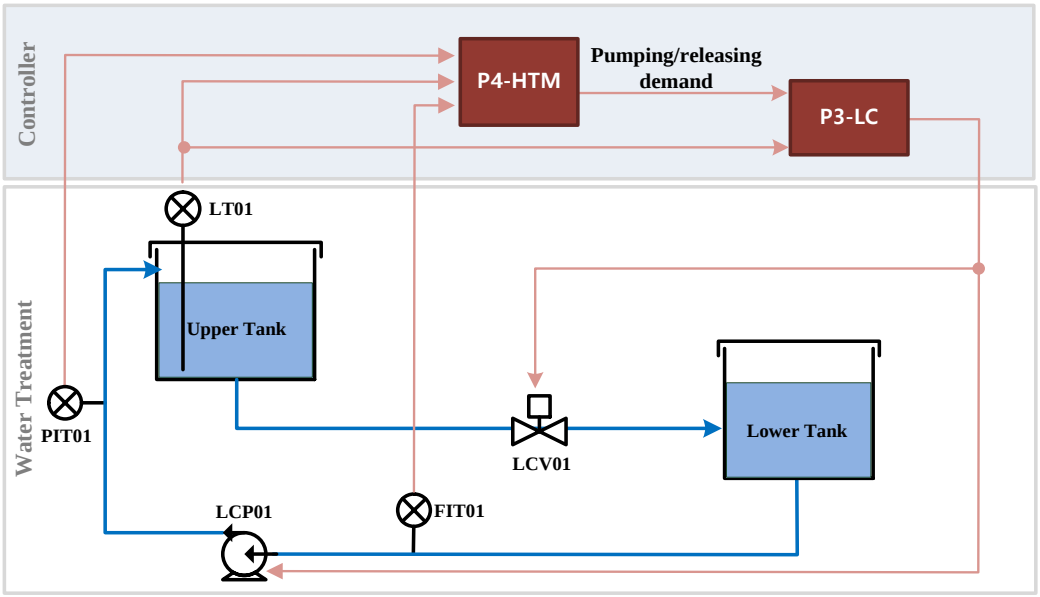

FIGURE 10. PROCESS ARCHITECTURE OF THE WATER-TREATMENT PLANT.

P3-LC: Level Control

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

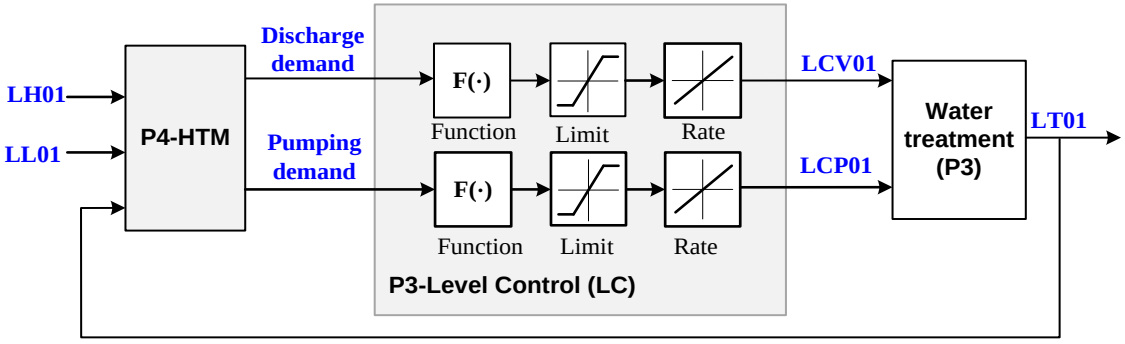

FIGURE 11. LEVEL CONTROL OF WATER-TREATMENT PLANT.

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
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 FCV01 valve	√	√
10	P1_FCV01Z	0	100	%	Current position of FCV01 valve	√	√
11	P1_FCV02D	0	100	%	Position command for FCV02 valve	√	√
12	P1_FCV02Z	0	100	%	Current position of FCV02 valve	√	√
13	P1_FCV03D	0	100	%	Position command for FCV03 valve	√	√
14	P1_FCV03Z	0	100	%	Current position of FCV03 valve	√	√
15	P1_FT01	0	2,500	mmH2O	Measured flowrate of return water tank	√	√
16	P1_FT01Z	0	3,190	l/h	Water inflow rate converted from P1_FT01	√	√
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 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 valve LCV01	√	√
22	P1_LCV01Z	0	100	%	Current position of valve LCV01	√	√
23	P1_LIT01	0	720	mm	Water level of return water tank	√	√
24	P1_PCV01D	0	100	%	Position command for valve PCV01	√	√
25	P1_PCV01Z	0	100	%	Current position of valve PCV01	√	√

No	Name	Range		Unit	Description	HAI	
		Min	Max			20.07	21.03
26	P1_PCV02D	0	100	%	Position command for valve PCV2	√	√
27	P1_PCV02Z	0	100	%	Current position of valve PCV02	√	√
28	P1_PIT01	0	10	bar	Heat-exchanger outlet pressure	√	√
29	P1_PIT02	0	10	bar	Water supply pressure of heating water pump	√	√
30	P1_PP01AD	0	1	Boolean	Start command of main water pump PP01A		√
31	P1_PP01AR	0	1	Boolean	Running state of main water pump PP01A		√
32	P1_PP01BD	0	1	Boolean	Start command of main water pump PP01B		√
33	P1_PP01BR	0	1	Boolean	Running state of main water pump PP01B		√
34	P1_PP02D	0	1	Boolean	Start command of heating water pump PP02		√
35	P1_PP02R	0	1	Boolean	Running state of heating water pump PP02		√
36	P1_STSP	0	1	Boolean	Start/stop command of boiler DCS		√
37	P1_TIT01	-50	150	℃	Heat-exchanger outlet temperature	√	√
38	P1_TIT02	-50	150	℃	Temperature of heating water tank	√	√
39	P2_24Vdc	0	30	Voltage	DCS 24V Input Voltage	√	√
40	P2_AutoGo	0	1	Boolean	Auto start button	√ (Auto)	√
41	P2_AutoSD	0	3,200	RPM	Auto speed demand	√ (SD01)	√
42	P2_Emerg	0	1	Boolean	Emergency button	√ (Emgy)	√
43	P2_ManualGo	0	1	Boolean	Manual start button		√
44	P2_ManualSD	0	3,200	RPM	Manual speed demand		√
45	P2_OnOff	0	1	Boolean	On/off switch of turbine DCS	√ (On)	√
46	P2_RTR	0	2,880	RPM	RPM trip rate		√
47	P2_SCO	0	100,000	-	Control output value of speed controller		√
48	P2_SCST	-100	100	RPM	Speed change proportional to frequency change of STM		√
49	P2_SIT01	0	3,200	RPM	Current turbine RPM measured by speed probe	√	√
50	P2_TripEx	0	1	Boolean	Trip emergency exit button	√	√
51	P2_VIBTR01	-10	10	㎛	Shaft-vibration-related Y-axis displacement near the 1st mass wheel	√ (VYT02)	√
52	P2_VIBTR02	-10	10	㎛	Shaft-vibration-related X-axis displacement near the 1st mass wheel	√ (VXT02)	√
53	P2_VIBTR03	-10	10	㎛	Shaft-vibration-related Y-axis displacement near the 2nd mass wheel	√ (VYT03)	√

No	Name	Range		Unit	Description	HAI	
		Min	Max			20.07	21.03
54	P2_VIBTR04	-10	10	μm	Shaft-vibration-related X-axis displacement near the 2nd mass wheel	√ (VXT03)	√
55	P2_VT01	11	12	rad/s	Phase lag signal of key phasor probe	√	√
56	P2_VTR01	-10	10	μm	Preset vibration limit for sensor P2_VIBTR01		√
57	P2_VTR02	-10	10	μm	Preset vibration limit for sensor P2_VIBTR02		√
58	P2_VTR03	-10	10	μm	Preset vibration limit for sensor P2_VIBTR03		√
59	P2_VTR04	-10	10	μm	Preset vibration limit for sensor P2_VIBTR03		√
60	P3_FIT01	0	27,648	-	Flow rate of water flowing into the upper water tank		√
61	P3_LCP01D	0	27,648	-	Speed command for pump LCP01	√	√
62	P3_LCV01D	0	27,648	-	Position command for valve LCV01	√	√
63	P3_LH01	0	70	%	High water level set-point	√	√
64	P3_LIT01	0	90	%	Water level of the upper water tank	√ (LT01)	√
65	P3_LL01	0	70	%	Low water level set-point	√	√
66	P3_PIT01	0	27,648	-	Pressure of water flowing into the upper water tank		√
67	P4_HT_FD	-0.02	0.02	mHz	Frequency deviation of HTM	√	√
68	P4_HT_LD	0	100	MW	Electrical load demand of HTM	√	√
69	P4_HT_PO	0	100	MW	Output power of HTM	√	√
70	P4_HT_PS	0	100	MW	Scheduled power demand of HTM	√	√
71	P4_LD	0	500	MW	Total electrical load demand	√	√
72	P4_ST_FD	-0.02	0.02	Hz	Frequency deviation of STM	√	√
73	P4_ST_GOV	0	27,648	-	Gate opening rate of STM		√
74	P4_ST_LD	0	500	MW	Electrical load demand of STM	√	√
75	P4_ST_PO	0	500	MW	Output power of STM	√	√
76	P4_ST_PS	0	500	MW	Scheduled power demand of STM	√	√
77	P4_ST_PT01	0	27,648	-	Digital value of steam pressure of STM	√	√
78	P4_ST_TT01	0	27,648	-	Digital value of steam temperature of STM	√	√
TOTAL						59	78

ATTACK OVERVIEW

SCENARIO CONFIGURATION

All attack scenarios are configured based on the four variables of the feedback control loop, namely the setpoints (SP), process variables (PV), control variables (CV), and control parameters (CP).

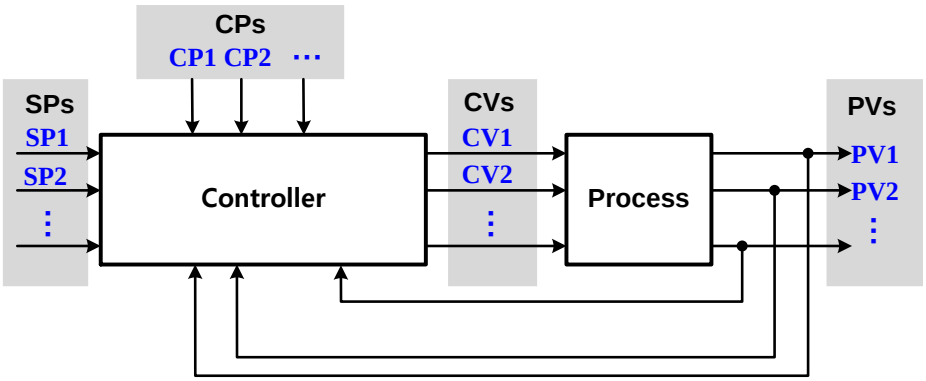

FIGURE 12. ATTACK MODEL BASED ON A PROCESS CONTROL LOOP.

NORMAL SITUATION

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.

A 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			
				Low Low	Low	High	High High
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 SCENARIO

Abnormal behavior 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
AP01	P1-PC	SP1	P1_B2016	Decrease 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 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	CV1	P1_PCV01D	Close pressure control valve of P1-PC. Restore to normal.	√	√
AP04	P1-PC	CV1	P1_PCV01D	Close pressure control valve of P1-PC. Restore to normal.	√	√
		PV1	P1_PIT01	Attempt to maintain previous sensor value.		
AP05	P1-PC	SP1-ST	P1_B2016	Short-term (ST) attack that decreases SP value of P1-PC for a few seconds and restores to normal. Repeat several times while hiding SP changes in HMI.		√
AP06	P1-FC	SP1	P1_B3005	Decrease SP value of P1-FC. Restore to normal while hiding SP changes in HMI.	√	√
AP07	P1-FC	SP1	P1_B3005	Decrease SP value of P1-FC. Restore to normal while hiding SP changes in HMI.	√	√
		PV1	P1_FT03	Attempt to maintain previous sensor value.		
AP08	P1-FC	CV1	P1_FCV03D	Open flow control valve of P1-FC. Restore in form of trapezoidal profile.		√
AP09	P1-FC	CV1	P1_FCV03D	Open flow control valve of P1-FC. Restore in form of trapezoidal profile.		√
		PV1	P1_FT03	Attempt to maintain previous sensor value.		
AP10	P1-FC	CV1-ST	P1_FCV03D	ST attack that opens flow control valve of P1-FC for a few seconds and restores to normal. Repeat several times.		√
AP11	P1-LC	SP1	P1_B3004	Increase SP value of P1-LC. Restore as a form of a trapezoidal profile while hiding SP changes in HMI.	√	√
AP12	P1-LC	SP1	P1_B3004	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.		
AP13	P1-LC	CV1	P1_LCV01D	Open level control valve of P1-LC and then restore in form of trapezoidal profile	√	√

Scenario	Target			Description	HAI	
	Controller	Variable	Point		20.07	21.03
AP14	P1-LC	CV1	P1_LCV01D	Open level control valve of P1-LC. Restore as a form of a trapezoidal profile.	√	√
		PV1	P1_LIT01	Open level control valve of P1-LC. Restore as a form of a trapezoidal profile.		
AP15	P1-LC	CV1-ST	P1_LCV01D	Attempt to repeat previous sensor value.		√
AP16	P2-SC	SP1	P2_AutoSD (P2_SD01)	ST attack that opens level control valve of P1-LC for a few seconds and restores to normal. Repeated several times.	√	√
AP17	P2-SC	SP1	P2_AutoSD (P2_SD01)	Decrease SP value of P2-SC. Restore to normal while hiding SP changes in HMI.	√	√
		PV1	P2_SIT01	Decrease SP value of P2-SC. Restore to normal while hiding SP changes in HMI.		
AP18	P2-SC	CV1	P2_SCO	Attempt to replay previous sensor value.		√
AP19	P2-SC	CV1	P2_SCO	Increase turbine control value of P2-SC. Restore to normal.		√
		PV1	P2_SIT01	Increase turbine control value of P2-SC. Restore to normal.		
AP20	P2-SC	SP1-ST	P2_AutoSD	Attempt to repeat previous sensor value.		√
AP21	P2-TC	SP1	P2_VTR01	ST attack that increases SP value of P2-SC for a few seconds and restore to normal. Repeated several times while hiding SP changes in HMI.		√
AP22	P2-TC	SP1	P2_VTR02	Open level control valve of P1-LC. Restore as a form of a trapezoidal profile.		√
AP23	P2-TC	SP1	P2_RTR	Open level control valve of P1-LC. Restore as a form of a trapezoidal profile.		√
AP24	P3-LC	CV1	P3_LCP01D	Attempt to repeat previous sensor value.		√
AP25	P3-LC	CV2	P3_LCV01D	ST attack that opens level control valve of P1-LC for a few seconds and restores to normal. Repeated several times.		√
AP26	P3-LC	SP1	P3_LH01	Decrease SP value of P2-SC. Restore to normal while hiding SP changes in HMI.	√	
		CV1	P3_LCP01D	Decrease SP value of P2-SC. Restore to normal while hiding SP changes in HMI.		
AP27	P3-LC	SP2	P3_LL01	Attempt to replay previous sensor value.	√	
		CV2	P3_LCV01D	Increase turbine control value of P2-SC. Restore to normal.		
TOTAL					14	25

DATASET

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.

An HMI operation task scheduler was used to periodically set 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	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)	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)

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,	15:47	83

No	ID	Attack			Start Time		Duration (sec)
		Scenario	Target Controller	Target Point(s)			
7	A202	AP02	P1-PC-SP1PV1	P1_B2016, P1_PIT01	2020	17:38	422
8	A203	AP15	P1-LC-CO1-ST7	P1_LCV01D		18:59	17
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 as “yyyy-MM-dd hh:mm:ss,” while the next 59 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.

The normal operations of the first training dataset (train1.csv) are given below, where all 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)

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
3	A103	AP14	P1-LC-CV1PV1	P1_LCV01D, P1_LIT01		15:45	868

No	ID	Attack			Start Time		Duration (sec)
		Scenario	Target Controller	Target Point(s)			
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
		AP07	P1-FC-SP1PV1	P1_B3005, P1_FT03			

No	ID	Attack			Start Time		Duration (sec)
		Scenario	Target Controller	Target Point(s)			
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		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	Nov. 5, 2019	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			

REFERENCES

DATASET

[20.08] 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.

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

PERFORMANCE METRICS

[19.11] Hwang, Won-Seok and Yun, Jeong-Han and Kim Jonguk and Kim Hyoungchun Kim, “Time-Series Aware Precision and Recall for Anomaly Detection: Considering Variety of Detection Result and Addressing Ambiguous Labeling”, CIKM ’19: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, pp.2241-2244, 2019.

- TaPR - https://github.com/saurf4ng/TapR

ABBREVIATIONS

C		
	CV	Control Variable
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	Level Switch [High/Low]
	LSL	Level Switch [Low]
	LT	Level 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