HAI SECURITY DATASET

HIL-based augmented ICS (HAI) security dataset was collected from a realistic industrial control system (ICS) testbed augmented with a hardwire-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 refining the dataset used in the competition. In 2022, HAI and HAIEnd 23.05 were developed for ICS Endpoint threat detection.

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 Vle 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.