SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

NATIONAL RESEARCH FOUNDATION
PRIME MINISTER'S OFFICE
SINGAPORE
Research . Innovation . Enterprise

iTrust
Centre for Research in Cyber Security

Dataset Characteristics

Secure Water Treatment (SWaT)

Characteristics of dataset (SWaT.A1_Dec 2015)

1. 11 days of continuous operation: 7 under normal operation and 4 days with attack scenarios
2. Collected network traffic & all the values obtained from all the 51 sensors and actuators
3. Data labelled according to normal and abnormal behaviours
4. Attack Scenarios: Derived through the attack models developed by our research team. The attack model considers the intent space of a CPS as an attack model. 41 attacks were launched during the 4 days and are described in the PDF.

Click here to find out how to request for the dataset.

Updates on dataset

24 Sep 18 (SWaT.A2_Dec 2015)
*Two sets of "SWaT_Dataset_Normal" – **versions 0 and 1** – are provided. The datasets capture the normal state of the SWaT testbed running for seven days. In **Version 0**, we started recording the data when the plant was emptying the water storage tank for 30 minutes. In general, in an ICS environment, this is part of the maintenance outside normal operations. As a result of this drainage, the first 30 minutes of LIT101 data exhibits change even though there was no water in/outflow. **Version 1** is derived from version 0 by removing the first 30 minutes of data.*

SWaT.A3_Jun 2017
136 hours of network traffic and historian data from continuously running SWaT (no attacks) was collected over 6 days.

14 Aug 19 (SWaT.A4_Jul 2019)
A new set of SWaT dataset, collected during Jul 2019, is available for downloading. This set includes 3 hours of SWaT running under normal operating condition and 1 hour during which 6 attacks were carried out. Those who have previously received SWaT dataset download link can use the same link to access this new dataset.

23 Oct 19 (SWaT.A5_Jul 2019)
We received queries on the SWaT dataset, collected during Jul 2019, e.g., under LS 201 the fields were recorded as "{u'IsSystem': False, u'Name': u'Inactive', u'Value': 0}". These fields have been updated to "Active" or "Inactive" and the dataset saved as version 2. The fields' definitions are provided in the "readme.docx" document that was shared along with the dataset. Those who were given the link to download the dataset previously can download the new files using the same link.

19 Dec 19 (SWaT.A6_Dec 2019)
A new set of SWaT dataset, collected during Dec 2019, is available for downloading. The

Critical Infrastructure Security Showdown (CISS)

CISS, which was originally named the SUTD Security Showdown (S3), has enabled researchers and practitioners to assess the effectiveness of methods and products aimed at detecting cyber attacks launched in real-time on SWaT.

S317
Data was collected during the 2017 run of S3, named S317. More details of S317 can be found here.

Characteristics of dataset

1. Network 'pcap' files for three days
2. Historian data for three days
3. Attack scenarios performed by the participants

Click here to find out how to request for the dataset.

CISS 2019

CISS2019.A1

On 27-30 Aug 2019, iTrust conducted the annual Critical Infrastructure Security Showdown (CISS 2019) exercise in SUTD. 4-days of data were collected during CISS 2019 from the SWaT testbed. On each day, the plant was running for 8 hours, from 0900 – 1300hrs and 1400 – 1800hrs (on the last day, the plant was only running from 0900 – 1300hrs). During this time, red teams launched attacks on the SWaT testbed. The CISS 2019 data consists of a total of 136,805 rows of state information. Each row contains measurements from 28 sensors as recorded in the SWaT Historian sampled at 1-second intervals. Two sets of data are available, one with attack information and one without.

CISS 2020-OL

The time-stamped dataset, containing one row every second, consists of two sets of Excel files. One set of files is labelled "Target-x" and the other as "CISS2020_OL-y." There are three target files and 18 CISS_OL files. Each Target file contains approximately one hour of data co while running SWaT under normal operating condition Each CISS_OL file contains data collected during

dataset consists of pcap and Historian Data (.csv) files. The dataset records a series of malware infection attacks on the SWaT Engineering Workstation. The malware attacks include Historian Data Exfiltration attack and Process Disruption attacks.

This set includes 3 hours of SWaT running under normal operating condition and 1 hour in which 6 attacks were carried out. Those who have previously received SWaT dataset download link can use the same link to access this new dataset.

9 Jul 2020 (SWaT.A7_June 2020)
SWaT was run on 4 occasions (no attack). Each run lasted either 2 or 4 hours. Network traffic data was captured for the 4 runs.

approximately a 4-hour run during which a red team launched attacks on SWaT.

Each Target data file contains 82 columns where each column corresponds to one state variable of SWaT. Each CISS2020_OL data file contains 97 columns. In addition to the SWaT state, as in Target files, the dataset contains attack information. Specifically, the following information is available: attack launch (AL), attack update (AU), attack target, attack type (IT, OT, or both), attack intent, attack mode, attack outcome (Success, Fail), attacker ID (anonymised), attack ID, and attack sub-ID.

Click here to find out how to request for the dataset.

Water Distribution (WADI)

Characteristics of dataset (WADI.A1)

1. 16 days of continuous operation: 14 under normal operation & 2 days with attack scenarios
2. Data from all the 123 sensors and actuators
3. Attack Scenarios: Derived from the attack models developed by our research team. The attack model considers the intent space of a CPS as an attack model. 15 attacks were launched during the 2 days.

Click here to find out how to request for the dataset.

Updates on dataset

19 Dec 19 (WADI.A2)
As the plant was unstable for certain periods during the operation, the affected readings have been removed and a new csv file "WADI_14days_new.csv" uploaded. A second csv file "WADI_attackdataLABLE.csv" now contains labels on whether there was an attack (-1) or not (1). The updated attack table with the corrected dates has also been uploaded.

BATtle of Attack Detection Algorithms (BATADAL)^

In addition to the WADI dataset, our faculty and researchers, in collaboration with colleagues from Israel and Cyprus, organised BATADAL, a competition to objectively compare the performance of algorithms for the detection of cyber attacks on water distribution systems. More information of the competition can be found here.

Characteristics of dataset

1. Training Dataset 1: This dataset was generated from a one-year long simulation. The dataset does not contain any attacks, i.e. all the data pertains to C-Town normal operations.
2. Training Dataset 2: This dataset is around 6 months long and contains several attacks, some of which are approximately labelled.

Click here to find out how to request for the dataset.

Electric Power and Intelligent Control (EPIC)

The dataset was collected by operating the EPIC testbed for 30 minutes under each of eight scenarios. Sensor measurements, actuator states and pcap files were collected.

Characteristics of dataset

Scenario 1

- Synchronization without load
- Angle difference between two generators G1 & G2 from -180 to 0 to 180 degree

Scenario 2

- Synchronization with 10kW resistive load
- Angle difference between two generators G1 & G2 from -180 to 0 degree

Scenario 3

- Two generators G1 & G2 running
- 10kW resistive load

Scenario 4

Scenario 5

- Two generators G1 & G2 running with PV system switched on
- 7kW resistive load

Scenario 6

- Three generators G1 to G3 running
- 14kW resistive load

Scenario 7

- Two generators G1 & G2 running
- Supplying power to iTrust's Secure Water Treatment (SWaT) testbed

Scenario 8

- Two generators G1 & G2 running with PV system switched on
- 10kW resistive load

- Two generators G1 & G2 running
- Supplying power to iTrust's SWaT and Water Distribution (WADI) testbeds

Click here to find out how to request for the dataset.

For publication on this dataset, please see below.

Blaq_0

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

Characteristics of dataset

Network 'pcap' files for three days were collected.

Click here to find out how to request for the dataset.

Internet of Things

Network traffic dataset collected by a high-interaction IoT honeypots deployed in the wild for 1.5 years during 2017-2018. The honeypots are manifested on 40 public IP addresses in the wild while forwarding the traffic to 11 real IoT devices. For more details, please refer to the publication "Scalable VPN-forwarded Honeypots: Dataset and Threat Intelligence Insights"

Click here to find out how to request for the dataset.

Complete Set of Invariants based on Design Centric and Data Centric Approaches

1. Set of Rules with antecedent 1
2. Set of Rules with antecedent 2
3. Set of Rules with antecedent 3
4. Set of Rules with antecedent 4
5. Set of Rules with antecedent 5
6. Set of Rules with antecedent 6
7. Set of Rules with antecedent 7
8. Comparison

Publication

[SWaT dataset] Goh J., Adepu S., Junejo K. N., and Mathur A., "A Dataset to Support Research in the Design of Secure Water Treatment Systems," The 11th International Conference on Critical Information Infrastructures Security.

[EPIC dataset] A Comprehensive Dataset from a Smart Grid Testbed for Machine Learning based CPS Security Research

^ If you are using the BATADAL datasets in your work, please cite the following paper as reference:

Riccardo Taormina and Stefano Galelli and Nils Ole Tippenhauer and Elad Salomons and Avi Ostfeld and Demetrios G. Eliades and Mohsen Aghashahi and Raanju Sundararajan and Mohsen Pourahmadi and M. Katherine Banks and B. M. Brentan and Enrique Campbell and G. Lima and D. Manzi and D. Ayala-Cabrera and M. Herrera and I. Montalvo and J. Izquierdo and E. Luvizotto and Sarin E. Chandy and Amin Rasekh and Zachary A. Barker and Bruce Campbell and M. Ehsan Shafiee and Marcio Giacomoni and Nikolaos Gatsis and Ahmad Taha and Ahmed A. Abokifa and Kelsey Haddad and Cynthia S. Lo and Pratim Biswas and M. Fayzul K. Pasha and Bijay Kc and Saravanakumar Lakshmanan Somasundaram and Mashor Housh and Ziv Ohar; "The Battle Of The Attack Detection Algorithms: Disclosing Cyber Attacks On Water Distribution Networks." Journal of Water Resources Planning and Management, 144 (8), August 2018. (doi link, bib)

[IoT dataset] Yan Lin Aung, Hui Hui Tiang, Herman Wijaya, Martín Ochoa, and Jianying Zhou. 2020. Scalable VPN-forwarded Honeypots: Dataset and Threat Intelligence Insights.
In Sixth Annual Industrial Control System Security (ICSS) Workshop (ICSS 2020), December 8, 2020, Austin, TX, USA. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3442144.3442146

Dataset collection credits

The following personnel were responsible for the dataset collection:

1. SWaT – Sridhar Adepu, Kaung Myat Aung, Desmond Wan, Beebi Siti Salimah Binte Liyakkathali
2. WADI – Venkata Reddy
3. S317 – Nils Tippenhauer, Hamid Reza Ghaeini
4. EPIC – Ding Liqun, Kandasamy Nandha Kumar, Chuadhry Mujeeb Ahmed

5. BATADAL – Riccardo Taormina
6. Blaq_0 – Francisco Furtado, Lauren Goh, Jonathan Heng
7. CISS 2019 – Desmond Wan, Francisco Furtado
8. IoT – Yan Lin Aung
