DATASETS

Since 2020, four 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/HAIEnd 23.05

HAI 23.05 and HAIEnd 23.05 were collected at the same time. HAI 23.05 and HAIEnd 23.05 include four training datasets, two testing datasets, and one label dataset in the form of CSV file. The time-series data in each CSV file satisfies time continuity. The first column represents the observed time in the "yyyy-MM-dd hh:mm:ss" format, and the remaining columns provide the recorded SCADA data points. The label dataset was marked as 1 only when attack occurred to indicate the presence or absence of an attack.

NORMAL OPERATION

We 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.

0.5

init

1.0

S1

init

S2

P2-SC

0.7

0.1

SP: P2_AutoSD

0.9

1.0

Range: [100, 120)

0.1

0.1

1.0

S2

0,3

0.4

0.7

1.0

S3

0.9

1.0

P1-LC

S3

0.3

S1

SP: P1_B3004

P1-FC

P2-SC

Range: [400, 500]

1.0

SP: P1_B400B

0.5

0.5

SP: P2 ManualSD

Range: [950, 1050]

Range: [11, 12]

P1-PC

P1-TC

P1-CC

SP: P1_B2016

SP: P1_B4022

SP: P1_B2016

Range: [0.1, 0.2]

Range: [25,27]

Range: [26,28]

(a) HMM-P1

(b) HMM-P1&P2

0.5

init

0.8

S1

P4-STM

SP: P4_HT_PS

Range: [0, 15]

0.2

0.5

1.0

SP: P4_ST_PS

Range: [0,50]

S2

P4-HTM

(c) HMM-P4

FIGURE 15. HMM-BASED GENERATIVE MODELS FOR NORMAL OPERATION.

30