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 feedback control loops in three different types of industrial control devices, namely the Emerson Ovation, GE Mark-Vle, and Siemens S7-1500. From the version 23.05, we also provide HAIEnd dataset that includes more detailed information about the internal control logic behaviors for Emerson boiler process control.

Version History

Four major versions of HAI dataset 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:

Normal Dataset

Abnormal Dataset

Release

of

Version

tags

File

Duration

Size

File

of

Duration

Size

(CSV)

(hours)

(MB)

(CSV)

attack

(hours)

(MB)

end-train1

250.5

end-test1

48.2

78

14

15

hai-train1

154.9

hai-test1

29.8

end-train2

260.7

end-test2

204.8

81

38

64

hai-train2

161.3

hai-test2

126.8

HAIEnd 23.05

225

end-train3

112.7

HAI 23.05

86

35

hai-train3

69.4

end-train4

176,0

55

hai-train4

109.2

799,9

253,0

Sum

249

Sum

52

79

494.8

156.6

train1

26

50.7

test1

7

24

48.2

train2

56

108,9

test2

17

23

44,5

train3

35

66.7

test3

10

17.3

33,4

HAI 22.04

86

train4

24

45.7

test4

24

36

69.5

train5

66

125.6

train6

72

136,8

Sum

279

534,4

Sum

58

100.3

195.6

train1

60

110

test1

5

12

22

train2

63

116

test2

20

33

61

train3

229

245

test3

8

30

55

HAI 21.03

78

test4

5

11

20

test5

12

26

47

Sum

352

471

Sum

50

112

205

train1

86

127

test1

28

81

119

HAI 20.07

59

train2

91

98

test2

10

42

62

Sum

177

225

Sum

38

123

181

Note: 1) The version numbering follows a date-based scheme, where the version number indicates

the released

date of a HAI dataset. 2) HAI 23.05 has the same experimental configuration as that of 22.04, 3) Both HAI 23,05 and

HAIEnd 23,05 data were collected simultaneously in the same experiment.