SAND2021-8051D

The Center for Cyber Defenders
Expanding computer security knowledge

Detecting Anomalies in Industrial Control Systems

Danyelle Loffredo, University of New Mexico

Project Mentor: Ryan Adams , Org 5628

Problem Statement:

- Detecting intrusions in industrial control systems (ICS) that are masked as legitimate commands can be difficult to detect, but important to stop. Very few labelled datasets on such attacks are available to developers of analytics that could detect these attacks.

Objectives and Approach:

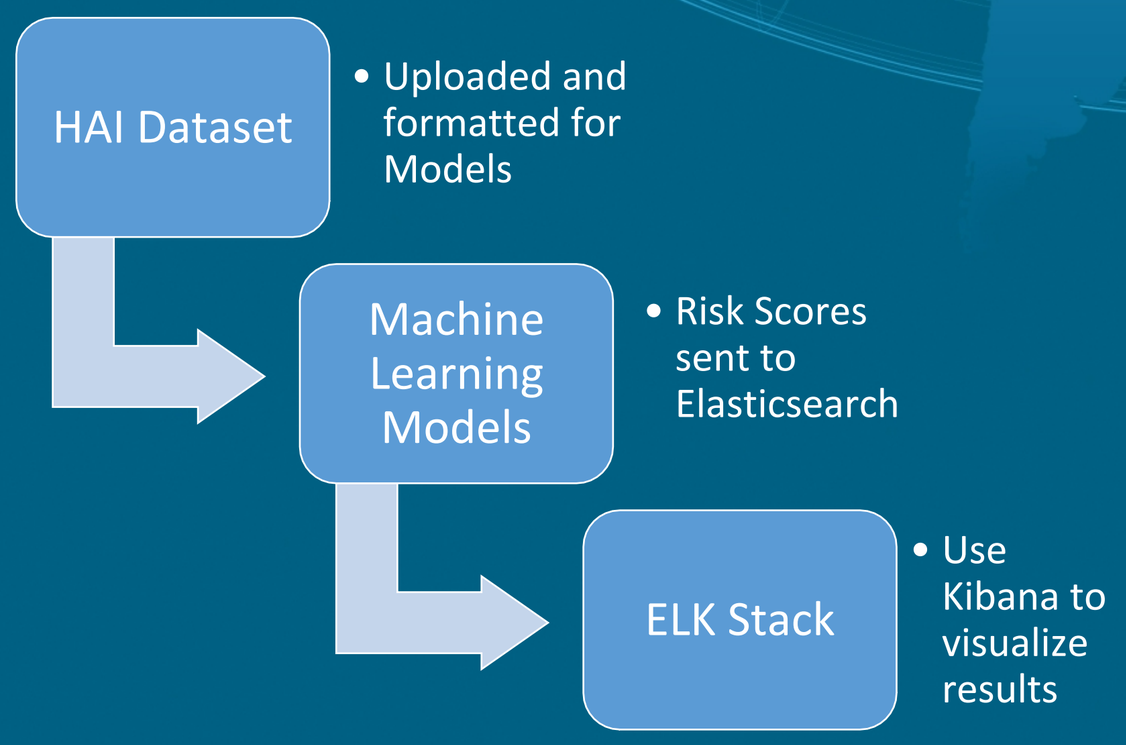
- Use machine learning techniques such as artis and principal component analysis (PCA) to analyze ICS data points for anomalies and process understanding
- Test the models on one of the few open source data sets, specifically the HAI dataset
- Present data in visualizations that organizations can use to take action against active events or to understand incidents in a forensics view

HAI Dataset Overview from Documentation

Version	Data Points	Normal Dataset			Attack Dataset			
		Files	Interval	Size	Files	Attack Count	Interval	size
HAI 21.03	78 points/sec	train1.csv	60 hours	100 MB	test1.csv	5 attacks	12 hours	22 MB
		train2.csv	63 hours	116 MB	test2.csv	20 attacks	33 hours	62 MB
		train3.csv	229 hours	246 MB	test3.csv	8 attacks	30 hours	56 MB
					test4.csv	5 attacks	11 hours	20 MB
					test5.csv	12 attacks	26 hours	48 MB
HAI 20.07 (HAI1.0)	59 points/sec	train1.csv	86 hours	127 MB	test1.csv	28 attacks	81 hours	119 MB
		train1.csv	91 hours	98 MB	test1.csv	10 attacks	42 hours	62 MB

https://github.com/icsdataset/hai

Process Breakdown



Results:

- Dataset dimensions were appropriately reduced using PCA
- Testing the models on the HAI dataset showed strong correlations between upticks in risk scores and confirmed attacks
- Results can be exported to Elasticsearch and viewed as visualizations in Kibana

Example dashboard from Kibana using HAI dataset

Impact and Benefits:

- Successful implementation of these techniques can lead to advancing the state and understanding of ICS analytics
- Advancements in ICS analytics using the HAI dataset could encourage organizations to create more opensource datasets

Sandia National Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. SAND2017-XXXXC.

U.S. DEPARTMENT OF ENERGY
National Nuclear Security Administration

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.