Project Idea Sketch

CS 6678 - Advanced Machine Learning
Instructor: Dr. Leslie Kerby

George Lake

Problem Area

The domain is Industrial Control Systems (ICS) / Critical Infrastructure Security

The ability to distinguish between benign sensor failures (e.g., wear and tear, calibration drift, etc.) and malicious cyber-attacks (e.g., sensor spoofing). Current anomaly detection systems often conflate the two, leading to high false alarm rates and operator desensitization.

I will classify three states: Normal, Physical Failure (sensor drift, bias, noise, etc.), and Cyber-Attack (spoofing, replay, etc.)

The dataset SWaT, labels the dataset as "Attacks" or "Normal", while the dataset Tennessee Eastman, labels the dataset as "Failures" or "Normal". My contribution will be modeling/injecting realistic sensor faults into the test set and seeing if a model trained on attacks can distinguish between them (or vice versa)

Potential Data

Here are two possible datasets. I will most likely use one of these, depending on how I want to go with the project. Either way, I will most likely be modifying them to add known events.

- iTrust - SWaT Dataset
  The Secure Water Treatment (SWaT) dataset is used for research in the cybersecurity of Industrial Control Systems and Cyber-Physical Systems. It was generated by the iTrust Center for Research in Cybersecurity at the Singapore University of Technology and Design. The dataset is derived from a fully operations, scaled-down water treatment plant. Normal data was collected over 7 days where the system operated under standard conditions without any interference. Attack data was collected over the final 4 days, during which many different cyber-attacks were launched against the system.
- Tennessee Eastman
  The Tennessee Eastman Process (TEP) dataset is used for research into industrial anomaly detection. This dataset is based on a chemical manufacturing process and was introduced by Downs and Vogel to provide a realistic simulation for process control and monitoring. The TEP simulates a plant where gaseous elements react to produce two liquid products. This dataset is a mathematical simulation, but is widely considered to be one of the best datasets for process control. The dataset contains many predefined faults which are the focus for anomaly detection.

Why This is Interesting

Practical Motivation: In an Industrial Control System, the physical manifestation of a sensor failure and a malicious cyber-attack can be very similar. However, the response from plant personnel is fundamentally different. Treating an attack as a fault allows an adversary to maintain persistence and escalate their attack. Conversely, treating every sensor drift as an attack leads to alarm fatigue, and costly operational shutdowns. Being able to differentiate between these events will be very useful.
Scientific Insight: Most current research in ML-based anomaly detection focuses on binary classification (normal vs. abnormal). This model could potentially differentiate between three different classifications (normal, fault, malicious)

Risks and Challenges

A sophisticated attacker might mimic a sensor failure (slowly drifting a value) to hide their activity. Distinguishing a "smart" attack from a "dumb" failure might be impossible without external ground truth, making evaluation difficult.