Scalable Industrial Control System Fuzzing Using Explainable AI

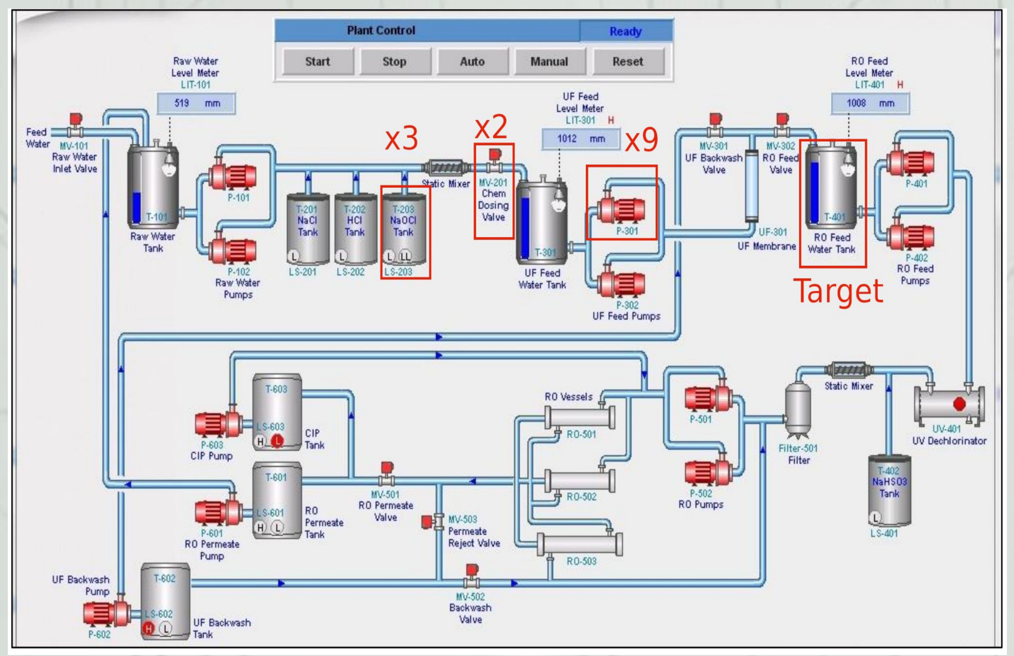
Justin Kur[1], Jingshu Chen[1] and Jun Huang[2]
1, Oakland University; 2, The City University of Hongkong

Overview

- Industrial Control Systems (ICS) are central to modern infrastructure, and their security is of paramount importance
- Fuzzing is one of most popular techniques to uncover failure-inducing input, but the large input space may make conventional fuzzing impractical
- We propose utilizing Explainable AI (XAI) for more sample efficient, understandable fuzzing

Case Study

An Overflow Water Tank



- Objective: Maximize water level sensor at the next time step
- Train attack model, and repeatedly sample actuator settings to learn feature importance
- Consistently set actuators are very important

Key Technical Points

- Represent each actuator as a high-dimensional embedding
  - So that the attack model is end-to-end differentiable, construct an attack as a linear combination of valid embedding vectors
  - Sample N(μ, σ) to get the contribution from each actuator setting
  - Scale vectors using softmax, and receive final embedding for attempted attack
- Learn optimal μ values for each actuator distribution
  - Fix σ, penalize high μ with L2 weight decay, so the attacker learns to control only important actuator values

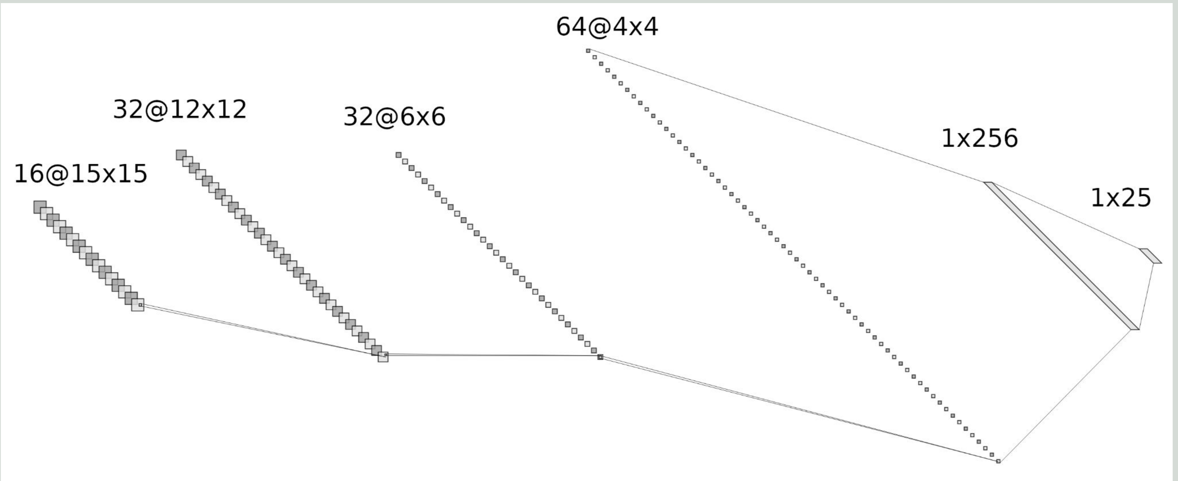
Our Approach

- Train a neural network on ICS dynamics with an operational trace
- Perform an attack on the neural network as a surrogate for the real system, without domain knowledge of that system
- Use XAI techniques to optimize an attack
  - Return importance of each actuator to the attack objective
  - Implicitly learn relevant dynamics of the ICS being modeled

System Modeling

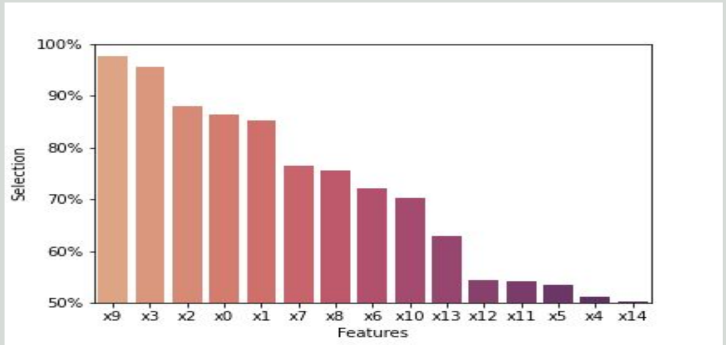
An Overflow Water Tank

- Convert actuators into embedding vectors, returning tensors of dimension (Time, Actuator,
- Two ResNet-style units (2D Convolution, ReLU, Average Pool, Batch Norm)
- concatenate ResNet output from actuators with sensor values and send through a final fully-connected layer



Results and Future Work

- Utilization of additional XAI techniques
- Unique attack for each initial condition
- Bootstrapping multiple step attacks
- Model limited control over actuators



THE TWENTY-THIRD ANNUAL
HIGH CONFIDENCE SOFTWARE AND SYSTEMS CONFERENCE
MAY 8 -10, 2023 | http://cps-hcss.org