Detecting Anomalies in Industrial Control Systems

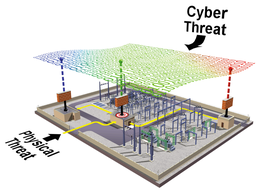
USING TIME-SAMPLED SENSOR AND ACTUATOR VALUES

Subin Sapkota, Dr. Hoda Mehrpouyan | DEPARTMENT OF COMPUTER SCIENCE, COLLEGE OF ENGINEERING, BOISE STATE UNIVERSITY

BOISE STATE UNIVERSITY

TOWARDS A RESILIENT AND SECURE CRITICAL INFRASTRUCTURE

INTRODUCTION



Industrial Control Systems (ICS) control physical processes in industries and critical infrastructure. They rely on sensors and actuators to collect the required information and manage the physical states of process.

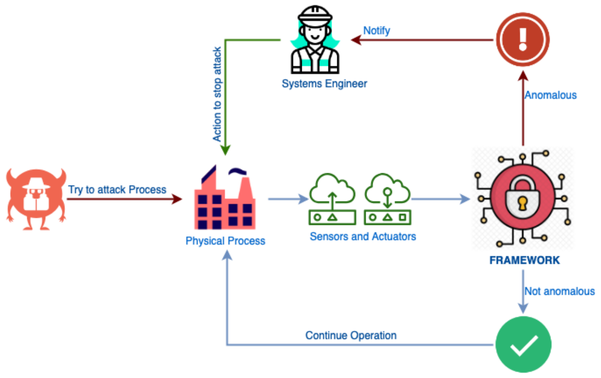
ICS are increasingly vulnerable to cyber-physical attacks, where cyber vulnerabilities are exploited to gain control of ICS to cause large environmental and monetary loss. To protect public facing critical infrastructures, ICS components must be protected from these new threats. While large number of research work in protecting infrastructure has been dedicated to Information Technology (IT) domain, few approaches have been solidified for ICS protection. Research in protection of ICS must be integral part of security for these systems.

Cyber-Physical Attacks

Cyber-physical attacks can take various forms, but in this paper we focus on detecting attacks that affect the sensors and actuators values. Attacks such as data manipulation of sensors values to affect actuation, false actuation, and delayed actuation can cause the physical process to deviate to unsafe conditions. For example, slowly changing the sensor value of a water Level Indicator to below lower bound value can cause controller to turn off pumps. This can be used to overflow tanks and cause damage to equipment. Catching these attacks require detection of sensor and actuator values deviation from normal behavior of system.

RESEARCH QUESTIONS

- In operational technology domain, how can we build a forecasting model to predict future behavior of sensor and actuator values based on previous behavior of these components?
- How can this forecasting model be used to predict anomalies that results from attacks, such as data injection and spoofing attacks, on sensors and actuators with high precision in real-time?
- Architecturally, these anomaly detection models can be trained over all subsystems at once, or distributed to sub-models for individual subsystems. Can we empirically determine the relative effectiveness of these different architectural designs using metrics such as anomalies detected, model size, training time, and detection time?



RELATED WORKS

Type	Contribution	Weakness
Machine Learning	Artificial Neural Network, as well as various classification algorithms were introduced.	Problem of interpretability of detected anomalies. Scalability is not tested. Number of new attacks caught is not specified, and large false positives were discovered.
Formal Models Specification	System model in modelling language to verify system safety and security properties.	Large amount of manual intervention required to model in formal language. Questions of scalability are not answered.

Anomaly Detection using Machine Learning and Deep Learning:

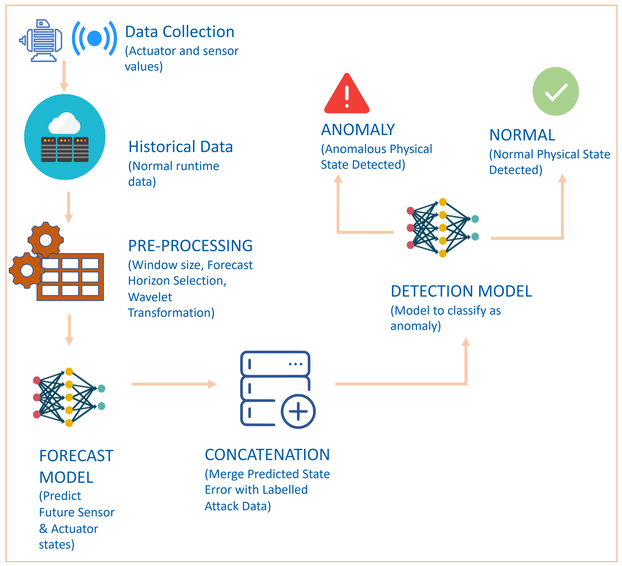
- Junejo, Khurum Nazir, and Jonathan Goh. "Behaviour-based attack detection and classification in cyber physical systems using machine learning." Proceedings of the 2nd ACM International Workshop on Cyber-Physical System Security. ACM, 2016.
- Shalyga, Dmitry & Filonov, Pavel & Lavrentyev, Andrey. (2018). Anomaly Detection for Water Treatment System based on Neural Network with Automatic Architecture Optimization.
- Goh, Jonathan, et al. "Anomaly detection in cyber physical systems using recurrent neural networks." 2017 IEEE 18th International Symposium on High Assurance Systems Engineering (HASE). IEEE, 2017.
- Inoue, Jun, et al. "Anomaly detection for a water treatment system using unsupervised machine learning." 2017 IEEE International Conference on Data Mining Workshops (ICDMW). IEEE, 2017.

Anomaly Detection using Formal Model Specifications

- Kang, Eunsuk, et al. "Model-based security analysis of a water treatment system." Proceedings of the 2nd International Workshop on Software Engineering for Smart Cyber-Physical Systems. ACM, 2016.
- Fauri, Davide, et al. "From system specification to anomaly detection (and back)." Proceedings of the 2017 Workshop on Cyber-Physical Systems Security and PrivaCy. ACM, 2017.
- Fauri, Davide, et al. "From system specification to anomaly detection (and back)." Proceedings of the 2017 Workshop on Cyber-Physical Systems Security and PrivaCy. ACM, 2017.
- Adepu, Sridhar, and Aditya Mathur. "Using process invariants to detect cyber attacks on a water treatment system." IFIP International Conference on ICT Systems Security and Privacy Protection. Springer, Cham, 2016.

FRAMEWORK OVERVIEW

Sensors and actuators data collected during normal system behavior of ICS are used to build a forecasting model to predict future values for sensors and actuators based on previous time window values. Wavelet transformation is applied to extract information from sensor values and add features. Error between predicted values and real values when the predicted time is reached is calculated. An anomaly detection model is trained by adding signals of attack on this error of predicted behavior to classify behaviors as anomalous/normal.



METHODOLOGY

- Sensor and actuator values are sampled every second during normal system runtime.
- Data is pre-processed to fit selected window-size and forecast horizon.
- Wavelet Transformation is used to decompose sensor signals to add features to each time window.
- Artificial Neural Network Model is built to forecast sensor and actuator values.
- Error of current state in terms of predicted state and actual states obtained.
- Error is concatenated to window-size data and labelled using the attack/normal signal from labelled attack dataset.
- Detection model that classifies a window as normal/anomalous is trained using Artificial Neural Network.

Case Study

Data

- Dataset obtained from a miniature water treatment plant: Secure Water Treatment System (SWaT) testbed.
- 52 sensor and actuator values sampled every second for 7 days normal behavior and 4 days attack data.
- Consists of 6 sub-processes.
- Combination of categorical and continuous values.

Attacks Analysis

- 36 attacks carried out with some form of impact.
- Targeting single points, single points in multiple processes, multiple points, or multiple points in multiple processes.
- Attacks take few minutes to 40+ minutes to make impact.

PRELIMINARY RESULTS

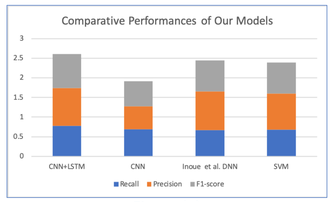


Figure (a)

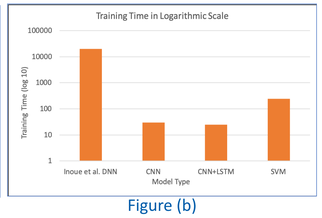


Figure (b)

Figures above show the comparative performance of our supervised model and the model presented in a state of the art paper using Deep Learning for anomaly detection in this domain (left figure). Our CNN+LSTM model misses 9 attacks occurring in the system, outperforming previous works that miss 23 attacks. We achieve this level of detection while exponentially decreasing the training times of our models (right figure). Currently, our model is dependent upon the signals provided by labeled attack data, and future work will be done to explore a semi-supervised level of detection models.

CONCLUSION

Preliminary results show ICS data based anomaly detection methods are promising and can be used to provide a new layer of security. Future work includes minimizing false alarms within the system and building larger labelled dataset of data that can bypass our anomaly detection system. This will help in continuous training of the framework to adapt to physical behavior they may misclassify. In addition, we will quantify detection time, training time, and model size for each variation of architecture explored using this methodology.