Cyber Risk Assessment of Traditional and Automated Machine Learning in Autonomous Control Systems

Patience Yockey

2023 LANNS Symposium

LANNS
LABORATORY FOR ADVANCED NUCLEAR NONPROLIFERATION AND SAFETY

Idaho National Laboratory

Idaho State University

iFAN Lab
Intelligence for Advanced Nuclear

Georgia Tech

Background & Purpose

LANNS
LABORATORY FOR ADVANCED NUCLEAR NONPROLIFERATION AND SAFETY

Background

- Advanced reactor designs include semi-autonomous or fully autonomous control systems (ACS).
    - Reduces overhead operations and maintenance cost [1].
    - Allows for online monitoring and diagnostics.
    - Makes minute adjustments to controls without human intervention.
- Machine learning (ML) based digital twins (DTs) are a growing consideration for ACS implementations.
    - Globally, the energy sector's use of ML is expected to grow by 29.88% between 2022 and 2029– equivalent to $37.4 billion [2].

Purpose

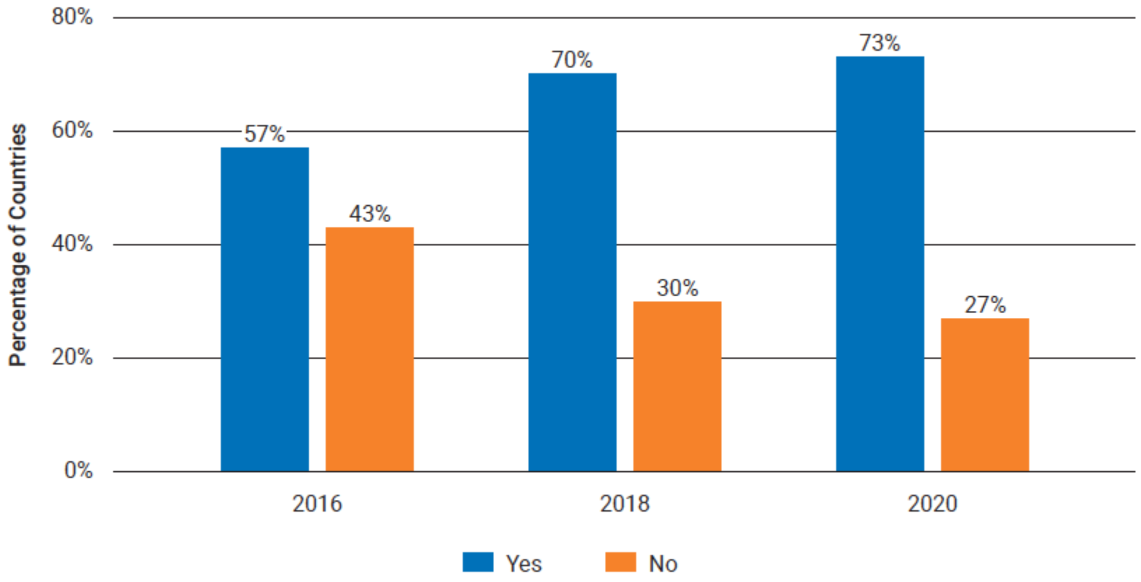
- This presentation is meant to highlight methodology and thought process behind determining cyber-risks and attack threats.
- More specifically how you (as nuclear engineers and medical physicists) can design systems with cybersecurity in mind.
- A simplified cyber-risk risk assessment of traditional and automated ML (AutoML) for ACS is shown in this presentation.

Georgia Tech

Motivation

LANNS
LABORATORY FOR ADVANCED NUCLEAR NONPROLIFERATION AND SAFETY

Percentage of Countries that Require Nuclear Facilities to Protect Against a Cyber Attack

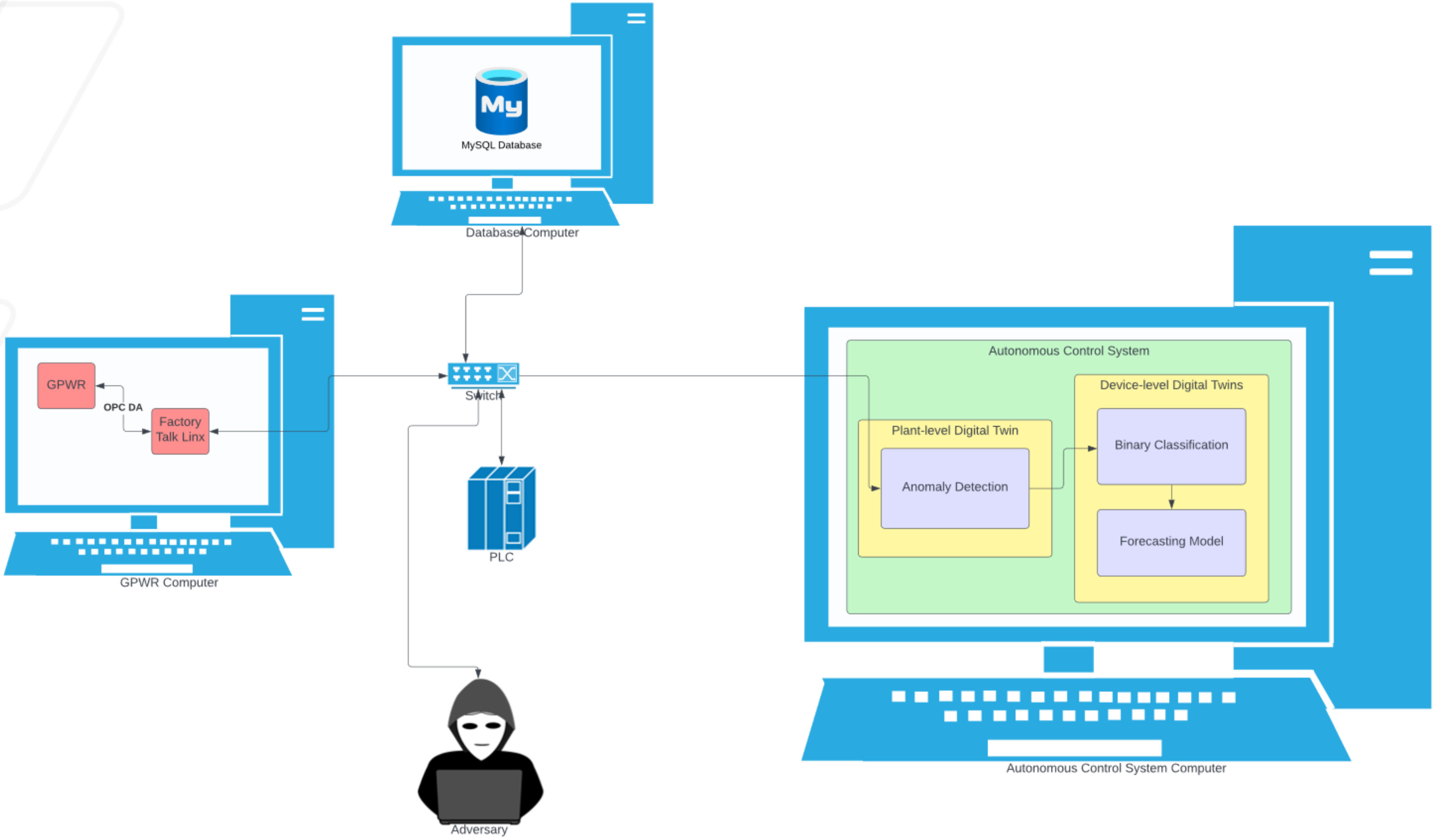


[3]

- **Cybersecurity** should be at the forefront of ACS development to ensure safe and reliable operations.
- According to the **Nuclear Security Index (NTI)**, cybersecurity is becoming more important for protecting nuclear facilities [3].
- **10 CFR 73.54:** ensure that “digital computer and communication systems and networks are adequately protected against cyber attacks, up to and including the design basis threat as described in § 73.1” [4].
- Previous work **by Idaho National Laboratory (INL)** includes a cyber-risk assessment framework for ACS [5].

Georgia Tech

Cyber-Physical Testbed Development



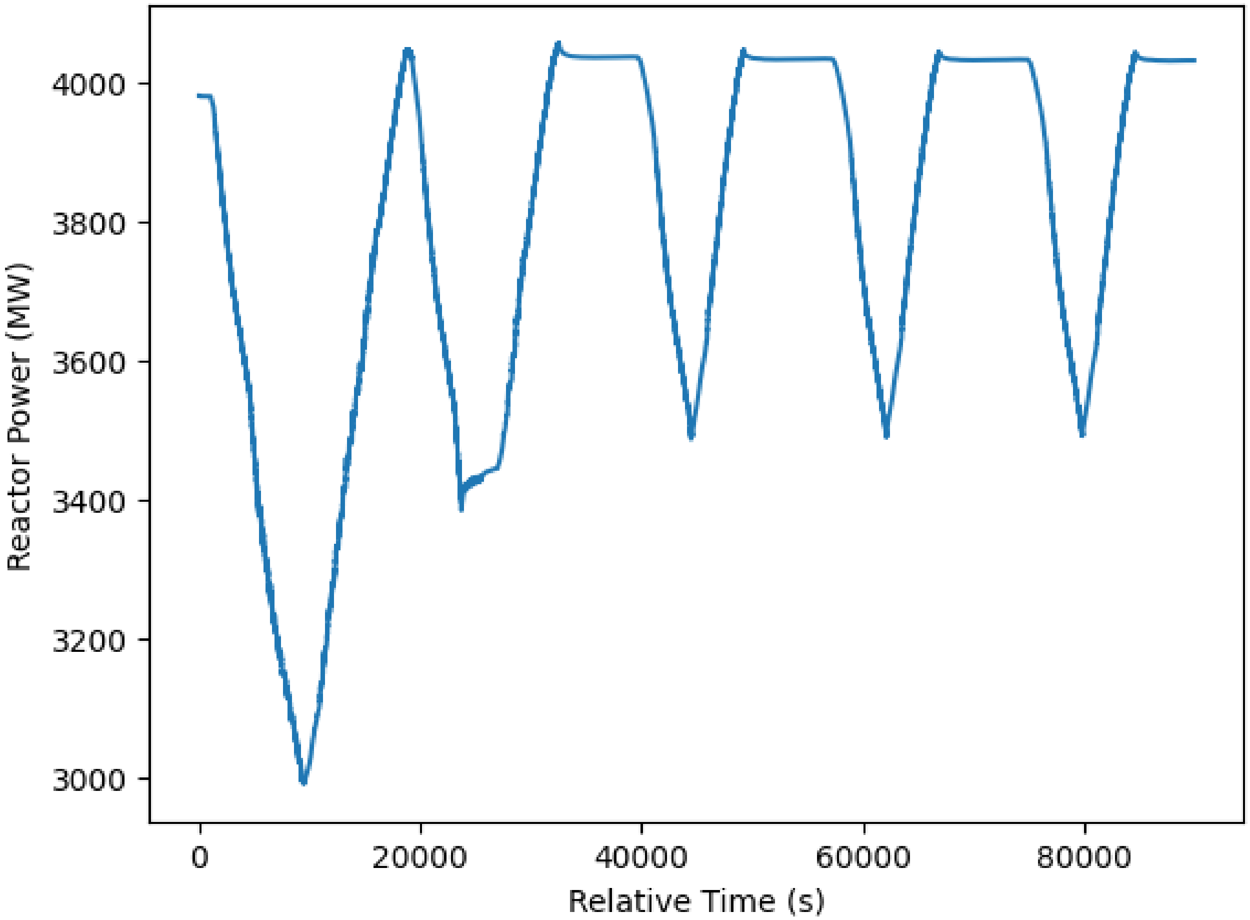
- **Cyber-Physical Testbed** includes the Generic Pressurized Water Reactor (GPWR) [6] and FactoryTalk Linx [7] to communicate to a programmable logic controller (PLC). The PLC communicates GPWR values to the ACS.
- **ACS** consists of a plant-level DTs to determine if the reactor is in an abnormal state and two device-level DTs to determine if the steam generator is undergoing a transient and forecasting steam generator flow rate.

In this scenario, the reactor is “air-gapped” from unsecured networks, meaning the adversary must have physical access to the system to launch attacks.

Data Collection and Storage

LANNS
LABORATORY FOR ADVANCED NUCLEAR NONPROLIFERATION AND SAFETY

Transient and Steady State Reactor Power Training Scenario



- On **training**, the ACS queries a separate system **MySQL database** [8] to obtain the training dataset for each DT.
  - Training data was obtained assuming a beginning of life (BOL) scenario.
    - Power was ramping over 420 minutes shown in figure.
    - The MySQL database pulled 70 variables related to steam generator 1 and overall plant-health ever 50 ms.
    - Training data was split into 70% training 30% validation.
- **Real-time data** is ingested by a **Pylogix** [9] call on the ACS following training to convert PLC CIP packets into usable dataframes.
  - Real-time data was 100% power BOL conditions to determine cyber-attack effects.

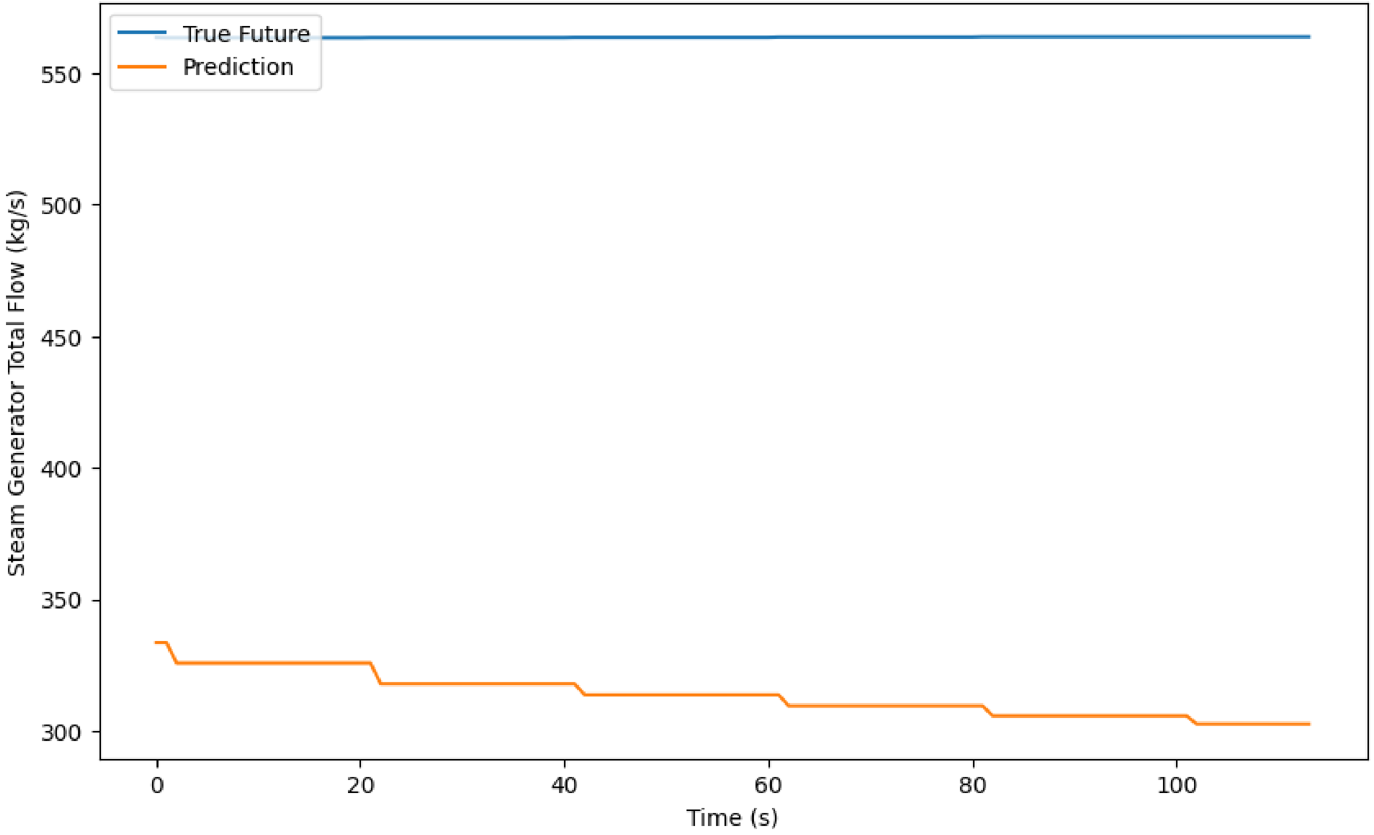
Georgia Tech

Unified Kill Chain

[10]

False Data Injections Attack Process

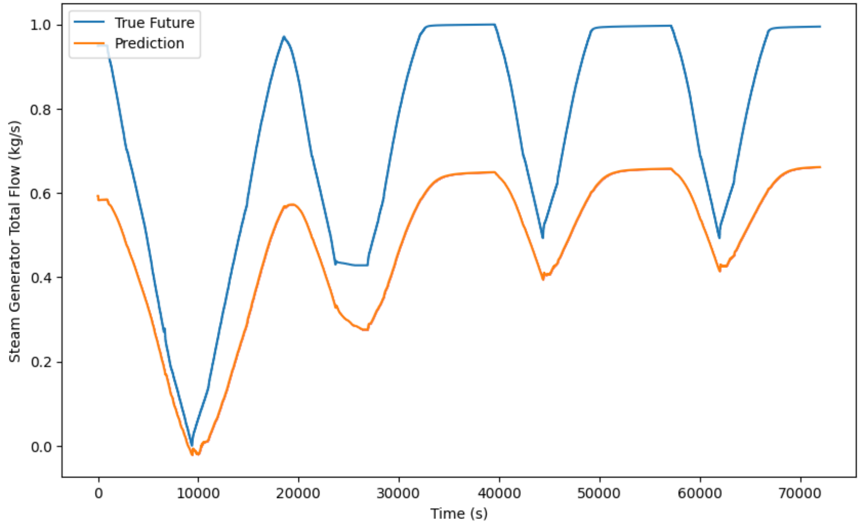
Traditional ML forecasted steam generator 1 flow after false data injection on real-time data



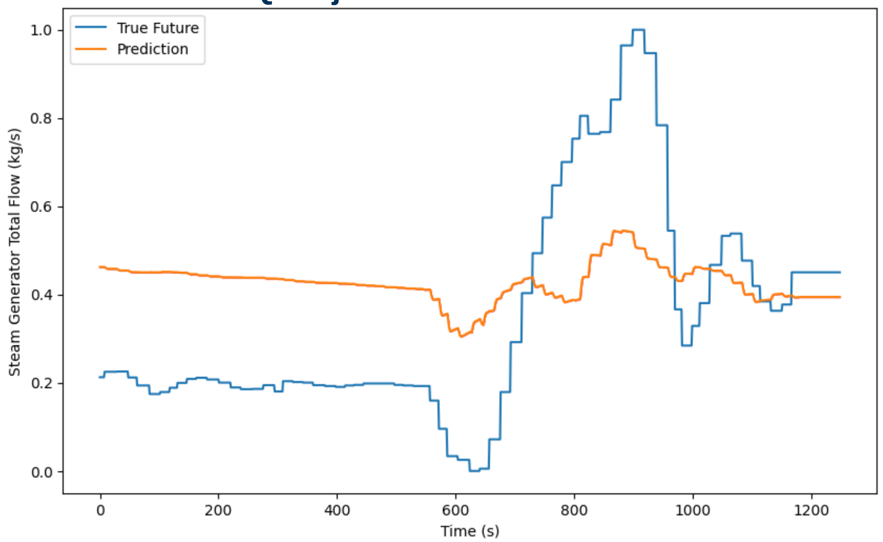
- PLCs use **Common Industrial Protocol (CIP)** [11] to communicate.
- **Initial Foothold:** Use Wireshark [12] while connected to internal network to filter out User Diagram Protocol (UDP) calls
    - Determine the PLC, GPWR, and ACS as well as PLC manufacturer.
- **Network Propagation:** Conduct a man-in-the-middle (MITM) attack using Ettercap [13] to address resolution protocol (ARP) poison the PLC and ACS.
- **Action on Objectives:** Collect and decrypt CIP packets to determine highly correlated GPWR tag values by looking up CIP tables.
    - Inject modified CIP packets using Scapy [14].
        - Steam generator 1 control valve positioning was reduced from nominal 50% open to 30% open.

SQL Injection Process

AutoML forecasted steam generator 1 flow after SQL injection on training data



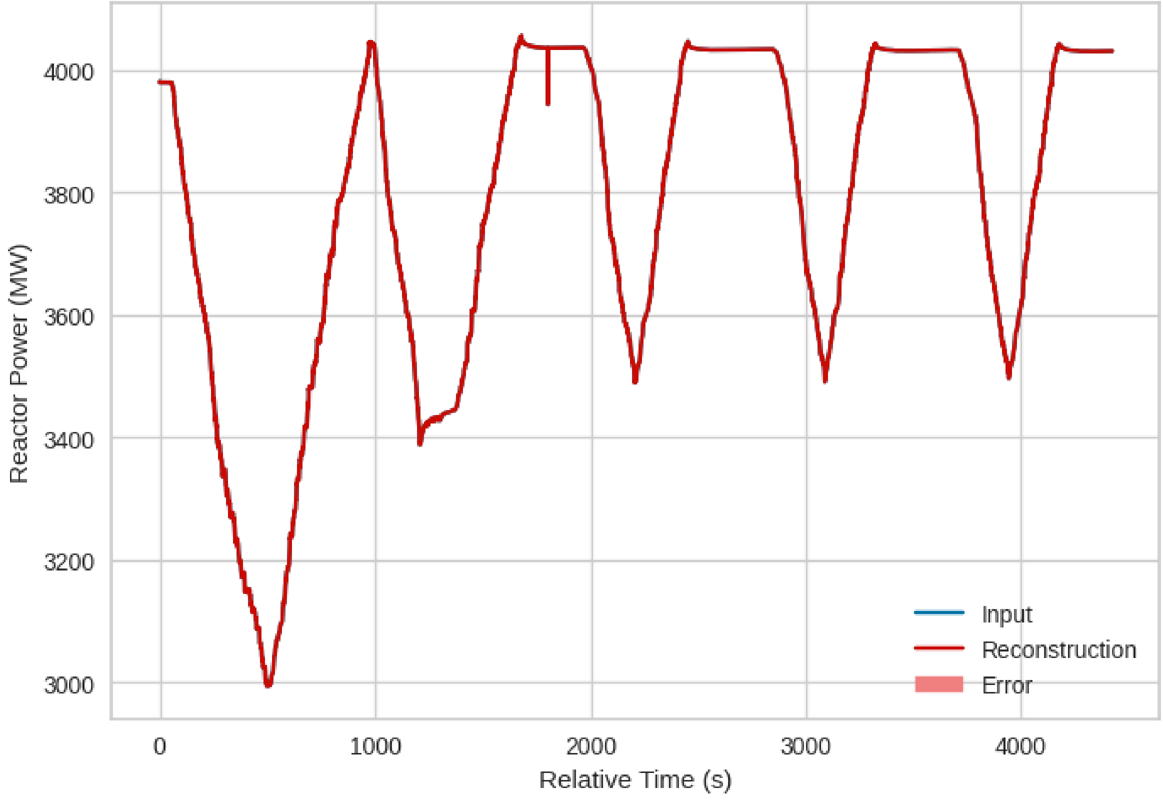
AutoML forecasted steam generator 1 flow after SQL injection on real-time data



- **Initial Foothold:** Use Wireshark to determine MySQL server by scanning for port 3306.
- **Network Propagation:** Use Metasploit [15] for a brute force dictionary attack to guess user password and login.
  - Otherwise, kill the server using a denial of service (DoS) attack using Hping3 [16], starting the server again using --skip-grant-tables to bypass authentication.
- **Action on Objectives:** Monitor SQL logs to determine what the autonomous control system is trained on.
  - Select and modify the highest correlated tags using SQL commands.
    - Steam generator 1 feedwater in was changed to allow be consistently the maximum flow rate.

Adversarial Hyperparameter/Weight Tuning

AutoML Reactor Power Training Reconstruction after Adversarial Hyperparameter Tuning



- Model **hyperparameters and weights** are typically saved periodically during the training process and can be accessed through the **Secure Shell protocol (SSH)** or **MySQL queries**.
- **Initial Foothold:** Use Wireshark to scan for port 22 for SSH or port 3306 for MySQL to determine the ACS.
- **Network Propagation:** Use Metasploit for a brute force dictionary attack on SSH or MySQL.
  - Otherwise, generate and send malicious code to user to add Secure Socket Layer (SSL) public keys for SSH or MySQL.
  - If choosing MySQL, a user defined function (UDF) will have to be created (if not preexisting) and executed to allow for Bash commands in MySQL.
- **Action on Objectives:** Decrypt files using the Python Joblib [17] library and modify the hyperparameters/weights.
  - Hyperparameters/weights were tuned to make 50% of the training dataset appear anomalous as opposed to the normal 1.7%.

Qualitative Cyber-Risk Assessment

LANNS
LABORATORY FOR ADVANCED NUCLEAR NONPROLIFERATION AND SAFETY

Traditional ML Cyber-Risk Matrix

Likelihood \ Impact	Low Impact	Medium Impact	High Impact
Low Likelihood	Hyperparameter Tuning via Malware	SQL Injections via authentication bypass	Hyperparameter /Weight Tuning via preexisting UDF
Medium Likelihood	-	Weight Tuning via Malware	-
High Likelihood	-	SQL Injections via Brute force Login	False Data Injections via CIP injections

AutoML Cyber-Risk Matrix

Likelihood \ Impact	Low Impact	Medium Impact	High Impact
Low Likelihood	Hyperparameter /Weight Tuning via preexisting UDF	Hyperparameter Tuning via Malware	SQL Injections via authentication bypass
Medium Likelihood	-	Weight Tuning via Malware	-
High Likelihood	False Data Injections via CIP injections	-	SQL Injections via Brute force Login

- Likelihood is qualified by complexity and amount of insider knowledge to complete the attack.
- Impact is qualified by the change in the model’s accuracy metrics following the attack.
- As shown, the likelihood of each of the attacks across both matrices stays the same, but the impact changes.

Georgia Tech

Risk Mitigation Strategies

- Mitigation strategies to reduce “red zone” risk
- Protecting ML models against SQL injections via brute force login:
  - Disable legacy authentication (password authentication) [18].
  - Set a maximum number of tries and sessions before being locked out of MySQL.
  - Set up a MySQL monitoring system to monitor for unauthorized or abnormal changes to training data.
- Protecting ML models against false data injections via CIP injections:
  - Implement a deep packet inspection (DPI) system and firewall for ethernet connections [19].
  - Use built in security functionality to encrypt CIP packets [20].

Summary & Future Work

Summary

- AutoML and traditional ML models are inherently tied for cyber-risk when examined in terms of likelihood and impact of false data injection, SQL injection, and adversarial hyperparameter/weight tuning.
- Traditional ML is more impacted by false data injections whereas AutoML is more impacted by SQL injections.
- To mitigate the risk of cyber attacks, strategies were presented for ACS system designers to implement during the development phase.

Future Work

- Develop more attack scenarios to fully encapsulate cyber-risk of traditional and AutoML.
- Determine if cyber-risk is different across different AutoML packages.
- Develop higher fidelity testbed to simulate more reactor subsystems and analyze the impact of cyber attacks on coupled devices.
- Implement and test the effects of proposed risk mitigation techniques.

Acknowledgements

Idaho National Laboratory

Idaho State University

iFAN Lab
Intelligence for Advanced Nuclear

References

[1] R. T. Wood, B. R. Upadhyaya and D. C. Floyd, "An Autonomous Control Framework for Advanced Reactors," Nuclear Engineering and Technology, vol. 49, no. 5, pp. 869-904, 2017.

[2] Maximize Market Research, "AI in Energy Market: AI is accelerating innovation in the energy sector as it continues to change the way of organizations and industries operate," December 2022. [Online]. Available: https://www.maximizemarketresearch.com/market-report/ai-in-energy-market/166396/#details. [Accessed 27 February 2023].

[3] Strengthen Cybersecurity at Nuclear Facilities - NTI Nuclear Security Index. (2020, July 22). NTI Nuclear Security Index. https://www.ntiindex.org/recommendation/recommendation-4-2/

[4] § 73.54 Protection of digital computer and communication systems and networks. (n.d.). NRC Web. https://www.nrc.gov/reading-rm/doc-collections/cfr/part073/part073-0054.html

[5] C. Spirito, P. Lamb, S. Aghara, C. Duffley, J. Strandburg, J. Coble and F. Zhang, "Cyber Threat Assessment for Autonomous and Remote Operations for Advanced Reactors," Idaho National Laboratory, Idaho Falls, 2021

[6] Western Services Corportation, "3KEYMASTER™ Generic Pressurized Water Reactor," Western Services Corportation.

[7] Rockwell Automation, "FactoryTalk Linx," Rockwell Automation, [Online]. Available: https://www.rockwellautomation.com/en-us/products/software/factorytalk/operationsuite/communications/linx.html. [Accessed 19 February 2023].

[8] Oracle, "MySQL," Oracle, 2023. [Online]. Available: https://www.mysql.com/. [Accessed 19 February 2023].

[9] D. Roeder, "Pylogix," GitHub, 2023. [Online]. Available: https://github.com/dmroeder/pylogix. [Accessed 19 February 2023].

[10] Pols, P. (n.d.). Unified Kill Chain: Raising Resilience Against Cyber Attacks. https://www.unifiedkillchain.com/

[11] Common Industrial Protocol (CIPTM) | ODVA Technologies. (n.d.). ODVA. https://www.odva.org/technology-standards/key-technologies/common-industrial-protocol-cip/

[12] Orebaugh, A., Ramirez, G., Beale, J., & Wright, J. D. (2007). Wireshark & Ethereal Network Protocol Analyzer Toolkit. https://ci.nii.ac.jp/ncid/BB01482189

[13] A. Ornaghi and M. Valleri, "Ettercap," [Online]. Available: https://www.ettercap-project.org/about.html. [Accessed 19 February 2023].

[14] S, R. R., Rohith, R., Moharir, M., & Shobha, G. (2018). SCAPY- A powerful interactive packet manipulation program. https://doi.org/10.1109/icnews.2018.8903954

[15] Maynor, D., & Wilhelm, T. (2007). Metasploit Toolkit for Penetration Testing, Exploit Development, and Vulnerability Research. In Elsevier eBooks. https://doi.org/10.1016/b978-1-59749-074-0.x5000-4

[16] Liang, L., Zheng, K., Sheng, Q., & Huang, X. (2016). A Denial of Service Attack Method for an IoT System. https://doi.org/10.1109/itme.2016.0087

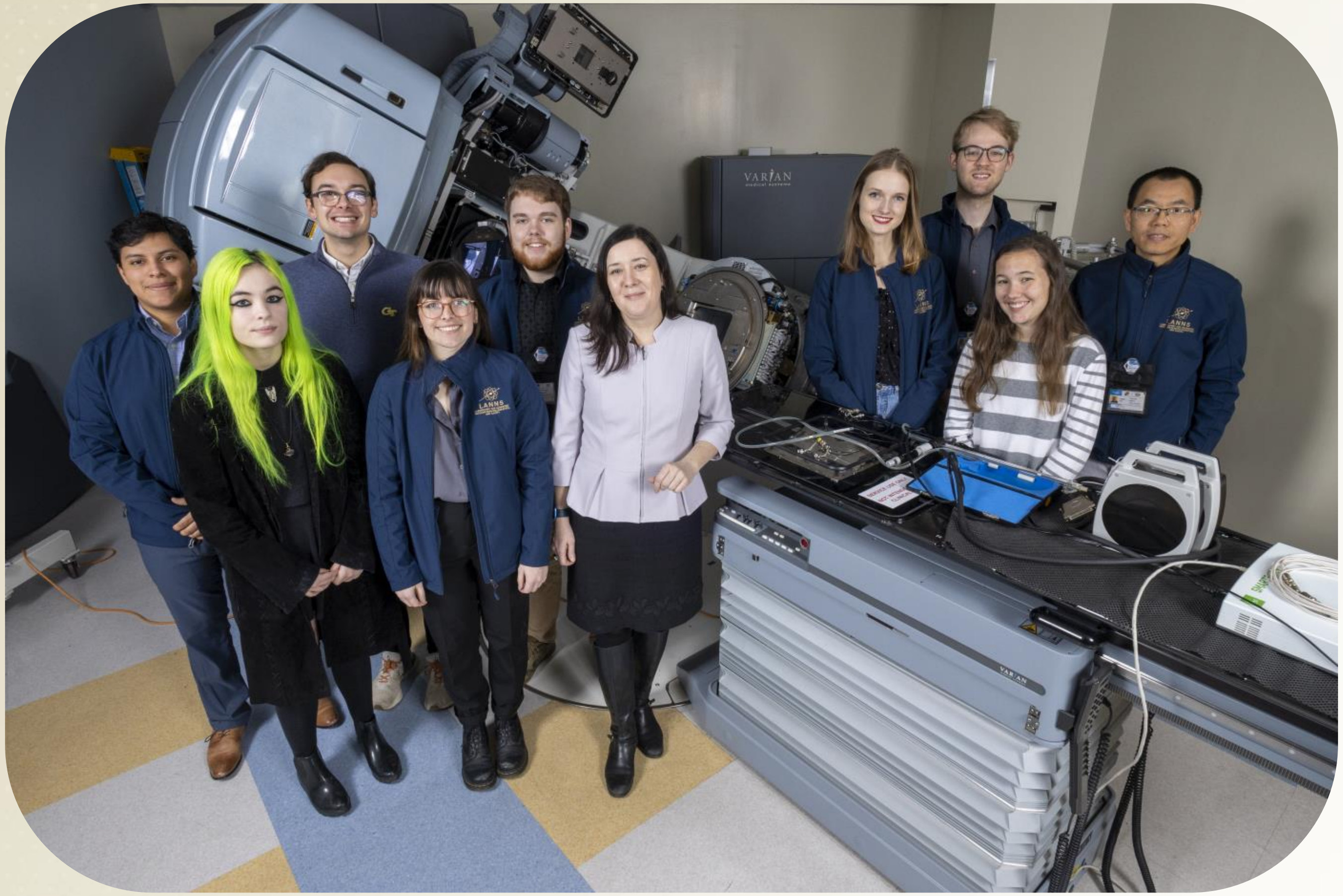
[17] PyCaret Team, "PyCaret Documentation," GitHub, 2023. [Online]. Available: https://github.com/pycaret/pycaret. [Accessed 19 February 2023].

[18] Microsoft, "Securing SQL Server," Microsoft, 17 February 2023. [Online]. Available: https://learn.microsoft.com/en-us/sql/relational-databases/security/securing-sql-server?view=sql-server-ver16. [Accessed 19 February 2023].

[19] Massachusetts Institute of Technology Review Insights, "Best practices for bolstering machine learning security," Massachusetts Institute of Technology, 14 November 2022. [Online]. Available: https://www.technologyreview.com/2022/11/14/1062881/best-practices-for-bolstering-machine-learning-security/. [Accessed 19 February 2023].

[20] Rockwell Automation, "ControlLogix 5580 and GuardLogix 5580 Controllers," Rockwell Automation, 2023. [Online]. Available: https://www.rockwellautomation.com/en-us/support/documentation/technical/programmable-controllers/controllogix-and-guardlogix-control-systems.html. [Accessed 19 February 2023].

Thank you



LANNS

LABORATORY FOR ADVANCED NUCLEAR NONPROLIFERATION AND SAFETY

Georgia Tech