INL/RPT-23-75941 Rev:000

U.S. DEPARTMENT OF ENERGY | Office of NUCLEAR ENERGY

CT-23IN110403: Analysis of Control System Ransomware Variants
CT-23IN110503: Evaluation of Advanced Sensors and Instrumentation
CT-23IN110501: Cyber Security for Nuclear Machine Learning Applications

Advanced Reactor Safeguards and Security

Chris Spirito
Idaho National Laboratory

CT-23IN110403: Analysis of Control System Ransomware Variants (i)

Overview

- Comprehensive exploration of ransomware threats to Industrial Control Systems (ICS) within the U.S. critical infrastructure.
- Emphasis on the deployment of virtual testbeds for organizations to simulate and test network vulnerabilities.
- Detailed methodologies for setting up virtual testbeds and conducting ransomware attack simulations.
- Exploration of various ransomware types, entry points, and potential mitigation strategies.

Why is this work important?

- Ransomware attacks on ICS pose significant threats to national security and critical infrastructure.
- The increasing frequency and impact of ransomware attacks necessitate proactive measures and robust cybersecurity practices.
- The ability to simulate and test vulnerabilities can help in the early detection and mitigation of potential threats.
- Understanding and categorizing Indicators of Compromise (IoCs) is crucial for safeguarding network and system security.

Why does industry care?

- Critical infrastructure sectors, including Energy and Government Facilities, have been targeted by ransomware attacks.
- Ransomware attacks can lead to significant operational, financial, and reputational damages.
- Virtual testbeds offer a cost-effective and adaptable solution for organizations to test and strengthen their cybersecurity defenses.
- The ability to simulate real-world scenarios can help organizations prepare for and respond to actual threats more effectively.

Key Results and Accomplishments

- Successful development and deployment of a modular virtual testbed for simulating ransomware attacks.
- Conducted experiments exploring diverse network entry points, ransomware architectures, and IoCs.
- Demonstrated the testbed's adaptability to distinct testing objectives, including the use of open-source features like ScadaBR.
- Provided comprehensive instructions for virtual testbed configuration and highlighted potential areas for further research and development.

CT-23IN110403: Analysis of Control System Ransomware Variants (ii)

Scenarios

- Attacker Profiles and Victim Profiles were crafted.
- Attacker Profiles included:
  - Access
  - Resources
  - Experience
- Victim Profiles included:
  - Sensitivity
  - Exposure
  - MITRE ICS ATT&CK Impact Codes
- Background
- Attack Sequence
- Experimental Format (goals, configuration, procedure, evaluation)
- Experimental Findings
- Knowledge, Skills, and Attitudes
  - Recommendations for ICT System Administrators
  - Recommendations for OT Personnel
  - Recommendations for Regulatory Authorities

Themes

- HMI Encryption and Exfiltration
- Altered Actuator State (Malicious State Command Injection)
- Altered Control Setpoint (Malicious Parameter Command Injection)

Character profiles

Attacker – Ransomware gang with limited ICS experience.

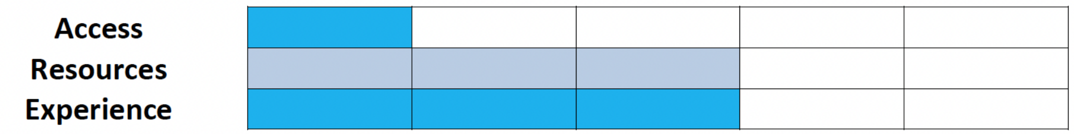


Figure 11: Scenario 1 Attacker Profile Table

Victim – Windows HMI connected to ICS system.

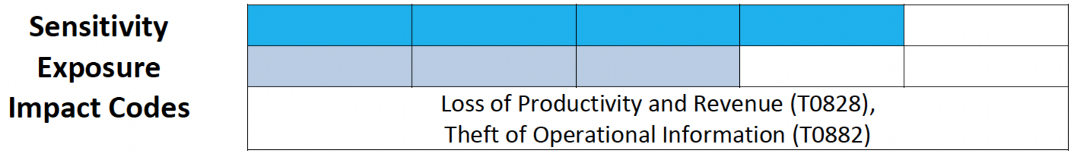


Figure 12: Scenario 1 Victim Profile Table

INL/RPT-23-75941 Rev:000

CT-23IN110503: Evaluation of Advanced Sensors and Instrumentation (i)

Overview

- Exploration of the integration and security of digital twins in nuclear power plants, with a focus on hyperparameter attacks on machine learning models.
- Detailed analysis of the potential vulnerabilities and threats posed by hyperparameter manipulations on digital twin machine learning models.
- Introduction of adaptive predictive control strategies and event-triggered mechanisms to counteract hyperparameter attacks.
- Comprehensive study of the implications of these attacks on the operational integrity and safety of nuclear power plants.

Why is this work important?

- The increasing integration of digital twins and machine learning models in nuclear power plants presents a new set of cybersecurity challenges.
- Hyperparameter attacks can compromise the predictive capabilities of digital twins, leading to potential operational inaccuracies and safety concerns.
- The nuclear industry's reliance on digital twins for optimizing operations and ensuring system reliability necessitates robust security measures.
- Ensuring the security and integrity of digital twins is paramount for the safety and reliability of nuclear power plants.

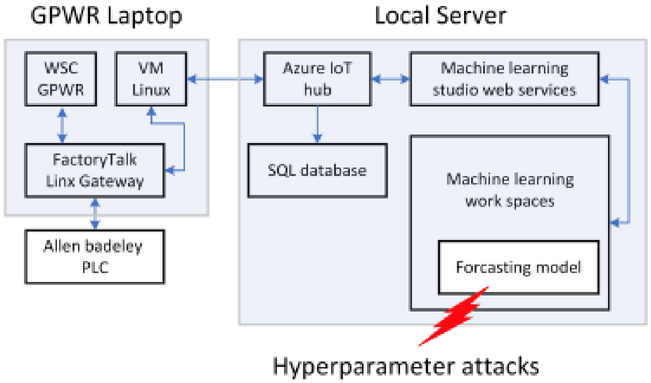
Why does industry care?

- Nuclear power plants are critical infrastructures, and any compromise in their operational integrity can have catastrophic consequences.
- The industry is progressively adopting digital twin technology, making it a prime target for cyber adversaries.
- Ensuring the security of digital twins can lead to cost savings, operational efficiencies, and enhanced safety protocols.
- Regulatory implications and the potential for stringent guidelines necessitate proactive measures to secure digital twin implementations.

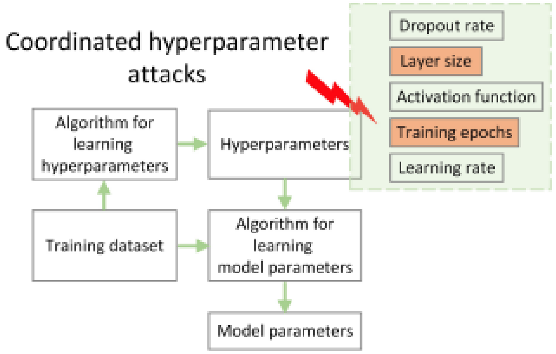
Key Results and Accomplishments

- Development and validation of an adaptive predictive control strategy anchored on an event-triggering law to counteract hyperparameter attacks.
- Successful simulation showcasing effectiveness and robustness of the proposed method against time-varying multi-rate hyperparameter attacks.
- Introduction of advanced control inputs, decision matrix integration, and machine learning-based predictive control for enhanced security.
- Recommendations for Regulators, Operators, and Cyber Teams to enhance security and resilience of digital twin implementations in NPPs.

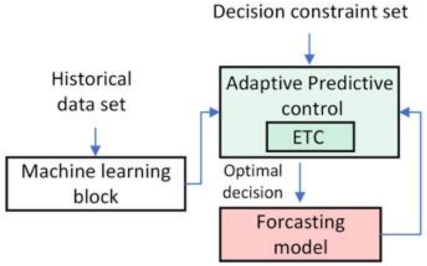
CT-23IN110503: Evaluation of Advanced Sensors and Instrumentation (ii)



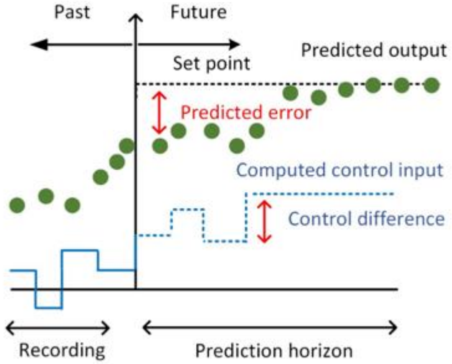
Hyperparameter attack on forecasting model of digital twin in NPP



Coordinated hyperparameter attack on ML forecasting model



Response Strategy Design



Adaptive Predictive Control Flow with event-triggered control

Algorithms

- Multi-Rate Time-Varying Coordinated Hyperparameter Attacks on Machine Learning Model
- Improved Predictive Control with Event-Triggered Adaptation

```
Algorithm 1: Multi-Rate Time-Varying Coordinated Hyperparameter Attacks on Machine Learning Model
Output: A decision matrix D_{i,j} to indicate the transition from a less intelligent attack to a high
        intelligent attack, and a function r(t) representing the multi-rate attack pattern over time
Initialize D_{i,j} ← zero matrix, attackSuccessful ← false, highIntelligenceAttack ← false,
  t ← 0, r(t) ← initial rate ;
if attackSuccessful = false then
  Penetrate multiple ML hyperparameters η, ρ, H, E, f(·); while true do
    if highIntelligenceAttack = false then
      // Low intelligence attack
      Randomly assign inappropriate values η′, ρ′, H′, E′, f′(·) based on multi-rate function
       g(t, r(t)); Apply changes to the model; Evaluate performance P′; if P′ < P − ΔP
       then
        highIntelligenceAttack ← true; D_{i,t} ← 1 for all affected hyperparameters i;
         break;
      end
      Update the attack rate r(t) based on predefined rule; t ← t + 1;
    end
    else if highIntelligenceAttack = true then
      // High intelligence attack
      Perform analysis of hyperparameters η, ρ, H, E, f(·); Identify and classify critical ones;
       Assign misleading values within threshold η″, ρ″, H″, E″, f″(·) based on multi-rate
       function h(t, r(t)) and optimal stealthy approach; Apply changes to the model;
       Evaluate performance P″; if P″ ≈ P − ΔP and P″ < P′ then
        attackSuccessful ← true; D_{i,t} ← 1 for all affected hyperparameters i; break;
      end
      Update the attack rate r(t) based on predefined rule; t ← t + 1;
    end
  end
end
return attackSuccessful, D_{i,j}, r(t);
```

```
Algorithm 2: Improved Predictive Control with Event-Triggered Adaptation
Input  : System matrices A, B, and data y; hyperparameters, control parameters μ, r, a, β, λ, and
         κ.
Output: Updated sliding window size, corrected hyperparameters, and optimal control gain
Step 1: Initialization;
Initialize variables for LMI, define variables P, K, M, N for optimization;
Step 2: Sliding Window and Hyperparameter Correction;
Function CalculateWindowSizePredictionHorizon()
  Init. ialize sliding window size and resize threshold;
  Initialize time-varying multi-rate function parameters; Initialize decision matrix D;
  foreach data point data in the window do
    Record data in the sliding window;
    if number of data points < 2 then
      Continue to the next data point;
    end
    Perform linear regression data window; Predict next data point; Calculate prediction error;
    if prediction error exceeds resize threshold then
      Increase the window size;
    end
    foreach hyperparameter h in hyperparameters do
      Calculate error err of the system with h;
      if err is too high then
        Determine intelligence level of the attack using decision matrix D;
        if attack is less intelligent then
          Adjust h using a lower intensity time-varying multi-rate function;
        end
        else
          Adjust h using a higher intensity time-varying multi-rate function;
        end
      end
    end
    foreach matrix M in matrices A and B do
      Evaluate model stability with M;
      if model is unstable then
        Adjust M to improve stability;
      end
    end
  end
Step 4: Algorithm Execution;
Call CalculateWindowSizePredictionHorizon();
Call ApplyPredictiveControlAndAdaptation();
Step 5: Return Results;
Return control gain K, updated hyperparameters, and system matrices;
```

M3CT-23IN1105033: Attack Surface Analysis of the Digital Twins interface with Advanced Sensor and Instrumentation

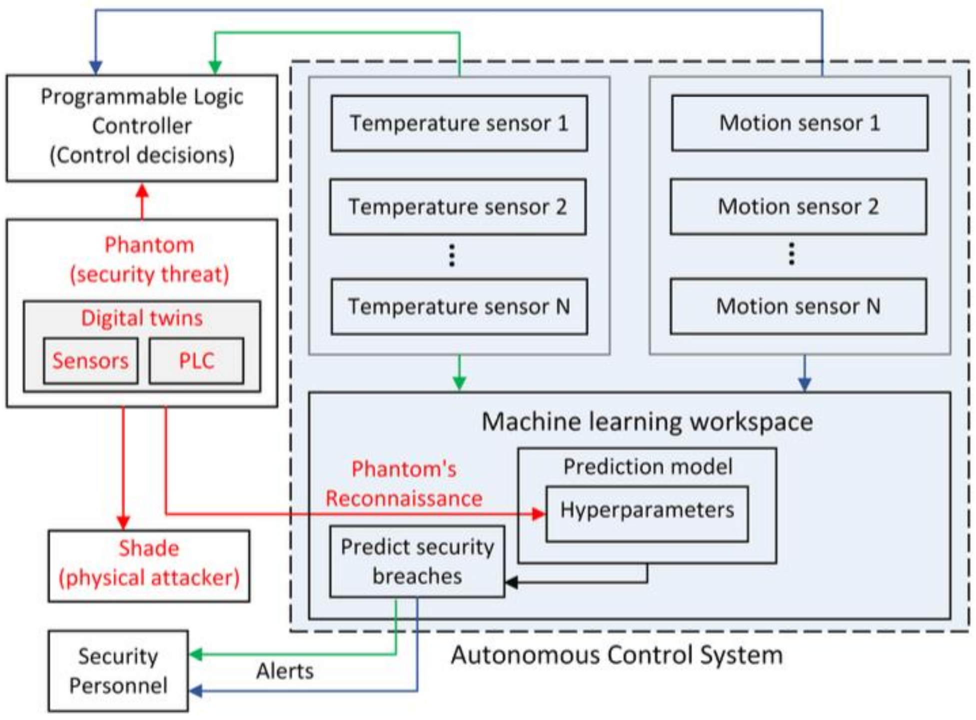
M3CT-23IN1105035: Protecting and Defending against Autonomous Control Systems and Digital Twins Cyber Attacks

INL/RPT-23-75941 Rev:000

CT-23IN110503: Evaluation of Advanced Sensors and Instrumentation (iii)

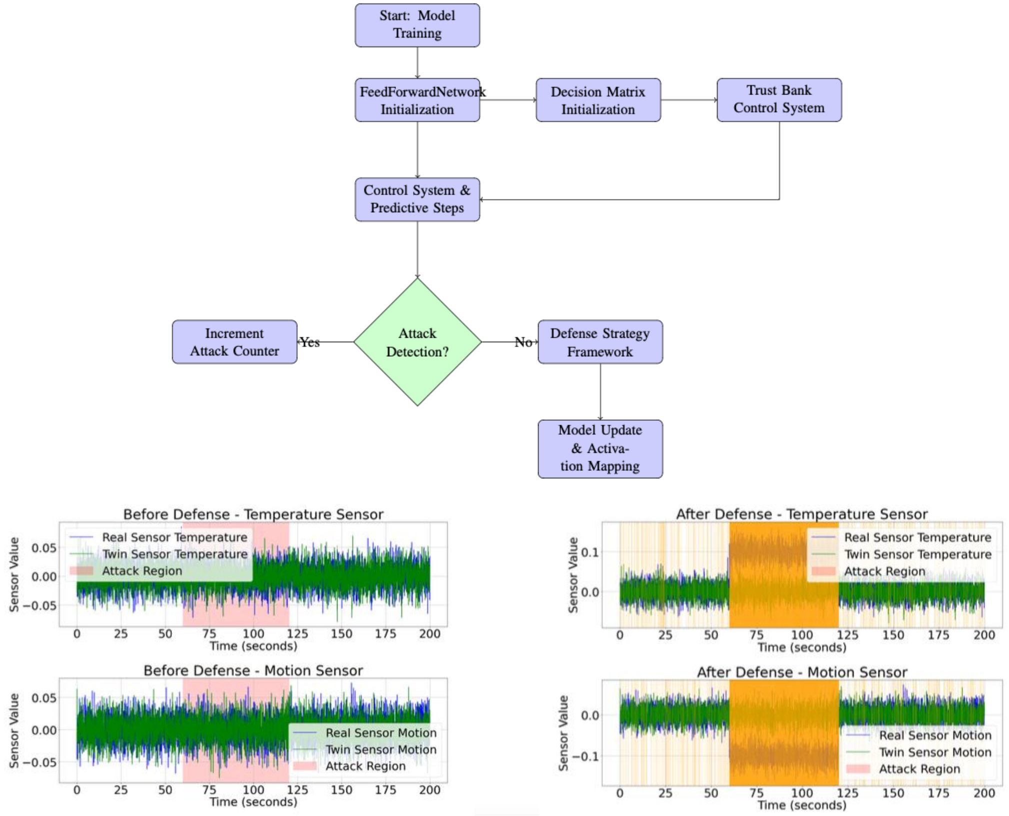
Operation FrostFire

Digital Twin-Assisted Hyperparameter Attack on Autonomous Control Systems



Operation MirrorShield

Proactive Defense against Hyperparameter Attack using Digital Twins



M3CT-23IN1105033: Attack Surface Analysis of the Digital Twins interface with Advanced Sensor and Instrumentation
M3CT-23IN1105035: Protecting and Defending against Autonomous Control Systems and Digital Twins Cyber Attacks

INL/RPT-23-75941 Rev:000

U.S. DEPARTMENT OF ENERGY | Office of NUCLEAR ENERGY

CT-23IN110503: Evaluation of Advanced Sensors and Instrumentation (iv)

Overview

- Comprehensive exploration of the digital twin technology, its applications in nuclear power generation, and its interface with Advanced Sensor and Instrumentation (ASI).
- Examination of potential security concerns, attack scenarios, and vulnerabilities associated with the deployment of digital twins in nuclear settings.
- Analysis of the unique requirements of digital twins and SMRs
- In-depth study of communication and data transfer interfacing between ASI and digital twin components, including protocols and specifications.

Why does industry care?

- The nuclear industry is progressively adopting digital twin technology, making it a prime target for cyber adversaries.
- Ensuring the security of digital twins can lead to cost savings, operational efficiencies, and enhanced safety protocols.
- The industry seeks to leverage the benefits of digital twins, such as predictive maintenance and streamlined operations, without compromising on security.
- Regulatory implications and potential legal liabilities necessitate proactive measures to secure digital twin implementations.

Why is this work important?

- Digital twins represent a transformative technology in the nuclear power industry, aiding in design, development, and predictive maintenance.
- The integration of digital twins in nuclear power plants introduces a new set of cybersecurity challenges and vulnerabilities.
- Ensuring the security and integrity of digital twins is paramount for the safety, reliability, and efficiency of nuclear power plants.
- The potential ramifications of a successful attack on a digital twin in a nuclear setting could be catastrophic, emphasizing the need for rigorous security measures.

Key Results and Accomplishments

- Identification and analysis of hypothetical attack scenarios, highlighting vulnerabilities in physical-digital sections of the digital twin infrastructure.
- Development of a baseline of recommended best practices for the deployment of digital twins in alignment with industry safety standards.
- Presentation of a POC, with documentation of methodology, simulations, and tests, showcasing real-life use cases and potential threats.
- Recommendations for security controls, compensating measures, and research pathways to enhance the security and resilience of DT in NPP.

M3CT-23IN1105033: Attack Surface Analysis of the Digital Twins interface with Advanced Sensor and Instrumentation
M3CT-23IN1105035: Protecting and Defending against Autonomous Control Systems and Digital Twins Cyber Attacks

U.S. DEPARTMENT OF ENERGY | Office of NUCLEAR ENERGY

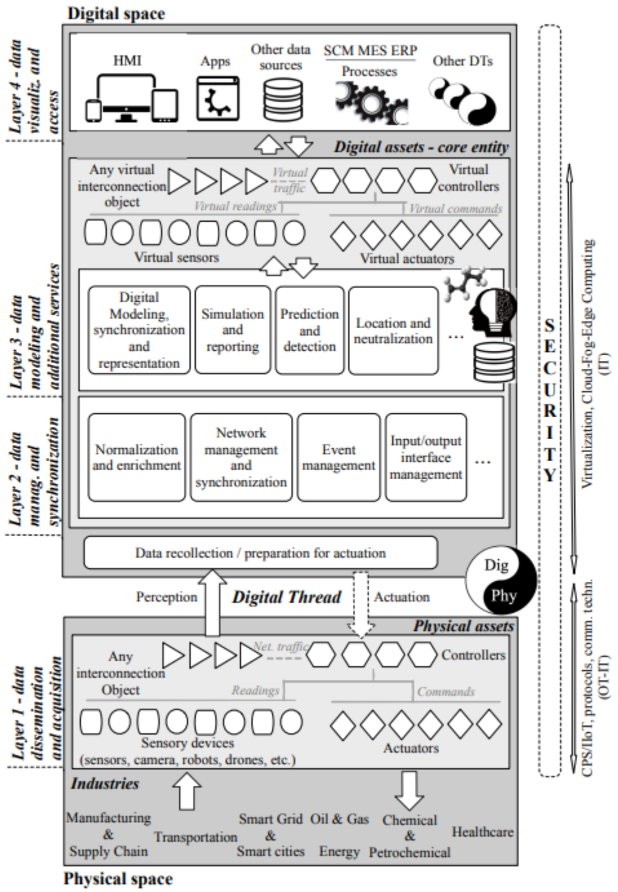
CT-23IN110503: Evaluation of Advanced Sensors and Instrumentation (v)

Overview

- Layer 1: Data Dissemination and Acquisition
- Layer 2: Data Management and Synchronization
- Layer 3: Data Modeling and Additional Services
- Layer 4: Data Visualization

Attack Scenarios

- Digital Twin Components
- Physical
  supply chain compromise
- Digital Section
  replication mode, simulation mode attacks
- Communication Section
  cloud computing interfaces



CT-23IN110501: Cyber Security for Nuclear Machine Learning Applications (i)

Overview

- Comprehensive exploration of the intersection between Artificial Intelligence (AI) and nuclear reactors, specifically focusing on their Instrumentation and Control (I&C) systems.
- Detailed guide tailored for three distinct audiences: I&C Vendors/Developers, Nuclear Regulators, and Nuclear Reactor Operators and Cyber Defense Teams.
- In-depth analysis of various cybersecurity threats associated with AI, including Inference Attacks, Adversarial Attacks, and Trojan Attacks.
- A structured approach to understanding the challenges and opportunities presented by Large Language Models (LLMs) in the context of nuclear reactor operations

Why does industry care?

- The nuclear industry is at the forefront of adopting advanced technologies, making it essential to understand and address the associated risks.
- Ensuring the security and integrity of AI implementations can lead to cost savings, operational efficiencies, and enhanced safety protocols.
- Regulatory implications, potential legal liabilities, and international standards necessitate proactive measures to secure AI implementations in nuclear reactors.
- The industry seeks to leverage the benefits of AI, such as predictive maintenance and streamlined operations, without compromising on security.

Why is this work important?

- AI, especially LLMs, is rapidly becoming integral to nuclear reactor operations, offering enhanced operational efficiency and anomaly detection.
- The integration of AI in nuclear reactors introduces a new set of cybersecurity vulnerabilities that need to be addressed proactively.
- Ensuring the secure and responsible implementation of AI in nuclear reactors is paramount for the safety, reliability, and efficiency of these critical infrastructures.
- As AI technologies evolve, understanding and mitigating associated threats is crucial to prevent potential catastrophic outcomes in the nuclear sector.

Key Results and Accomplishments

- Identification and detailed analysis of various AI-related cybersecurity threats, providing actionable insights and recommendations for mitigation.
- Development of a robust set of guidelines and recommendations tailored for I&C Vendors/Developers, ensuring secure development and deployment of AI.
- Comprehensive regulatory guidance for Nuclear Regulators, emphasizing the importance of updated guidelines, risk assessments, and adherence to international standards.
- Provision of technical reports, scenarios, and recommendations for Nuclear Reactor Operators and Cyber Defense Teams, ensuring the secure and optimized use of AI in operational environments.

M2CT-23IN1105017: Complete Evaluation on Advanced Reactor Machine Learning Subversion Attacks

M3CT-23IN1105013: Autonomous System Inference, Trojan, and Adversarial Reprogramming Attack and Defense

INL/RPT-23-75941 Rev:000

U.S. DEPARTMENT OF ENERGY | Office of NUCLEAR ENERGY

CT-23IN110501: Cyber Security for Nuclear Machine Learning Applications (ii)

Guide for Advanced Reactor & I&C Vendors

- **Recommendations for Developers (Technical Reports available)**
  - Inference Attacks
  - Large Language Model Attacks
  - Adversarial Attacks
  - Trojan Attacks
- **Scenarios**
  - Adversarial Attack (adversarial training)
  - Data Poisoning Attack

Guide for Nuclear Regulators

- **Recommendations for Regulators**
  - Guidelines for AI/ML inclusion in Cyber Security Plans
  - Legal Considerations
  - Risk Assessments and Human Studies

Guide for Operators and Cyber Defense Teams

- **Recommendations for Defenders**
  - Defensive Strategies for managing LLM use and access
  - Defensive Strategies for Large Language Model Attacks
    Protection against Membership Inference Attacks, Link Extraction Attacks,
  - Adversarial Attacks
    Identifying FGSM, Model Architectures
  - Defensive Strategies to protect against Trojan Attacks
    Data Management and Model Monitoring Recommendations, Use of Autoencoder Defenses, Neural Network Defenses
- **Scenarios**
  - Membership Inference
  - Trojan Attack
  - Use of Behavioral Analytics

M2CT-23IN1105017: Complete Evaluation on Advanced Reactor Machine Learning Subversion Attacks
M3CT-23IN1105013: Autonomous System Inference, Trojan, and Adversarial Reprogramming Attack and Defense
INL/RPT-23-75941 Rev:000

CT-23IN110501: Cyber Security for Nuclear Machine Learning Applications (iii)

Overview

- Comprehensive exploration of machine learning (ML)-driven autonomous control systems within advanced nuclear reactor designs.
- Detailed analysis of vulnerabilities and proposed strategies for defense against potential cyber-attacks, including Inference, Trojan, and Adversarial attacks.
- A crafted cyber-physical testbed and preliminary ACS were devised to reflect potential configurations of advanced reactor control designs.
- Recommendations and strategies are presented for both traditional and AutoML models, anchoring upon the existing knowledge landscape and ML-based DT modeling for ACS.

Why is this work important?

- The integration of autonomous control systems (ACS) within advanced nuclear reactor designs is becoming essential for operational efficiency.
- With a significant increase in cyber-attacks targeting the energy sector, there's a compelling necessity to fortify cybersecurity protocols in safeguarding reactor systems.
- The study extends beyond conventional cybersecurity parameters, diving into potential vulnerabilities woven into ML-based DTs and ACS in advanced reactor systems.
- Ensuring robust cybersecurity in this domain not only protects against immediate threats but also fortifies the infrastructure against evolving cyber challenges.

Why does industry care?

- Advanced cyber-attacks against critical infrastructure and the energy sector are becoming more common, necessitating robust defense mechanisms.
- The evolution towards advanced reactor systems utilizing digital instrumentation and controls (I&C) is essential for mitigating operations and maintenance costs.
- The integration of semi and fully autonomous control systems (ACS) emerges as a potent strategy to enhance economic feasibility of novel reactor designs.
- Ensuring the cybersecurity of ML-based DT technologies such as ACS prompts a holistic view of shared responsibility for maintaining cyber-secure ML-based systems.

Key Results and Accomplishments

- Development of defensive measures targeting protecting inference mechanisms and preventing unauthorized reprogramming.
- Strategies delineated for fortifying against Inference Attacks, thwarting Trojan Attacks, and mitigating Adversarial Attacks.
- Implementation of general defensive strategies, including maintaining offline backups, encrypting model data, employing multi-factor authentication, and engaging in regular audits.
- A multi-layered defense mechanism is orchestrated, ensuring comprehensive shielding against the spectrum of cyber threats in ML models integral to DT technologies and ACS implementations.

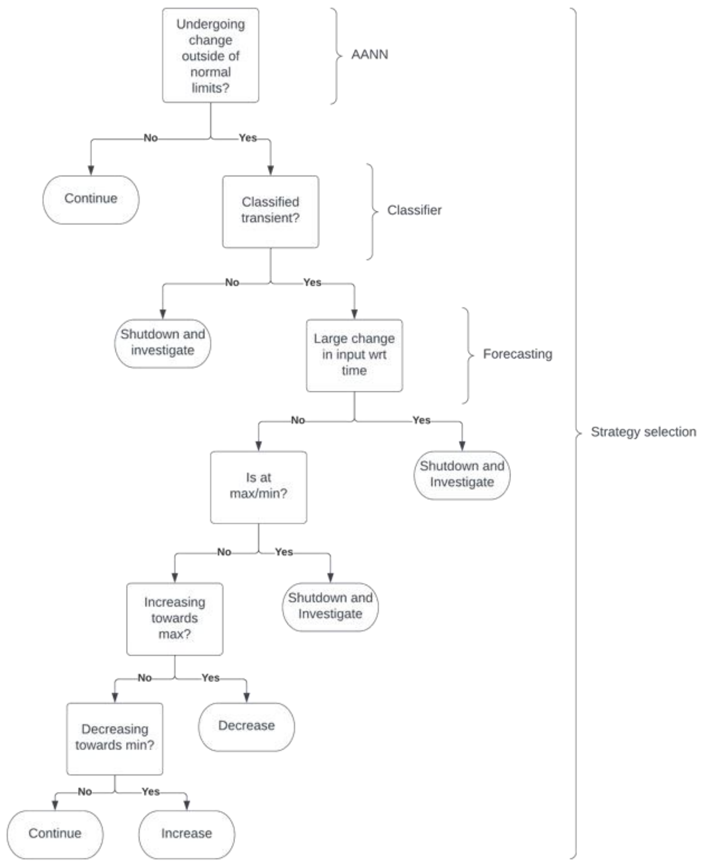
M2CT-23IN1105017: Complete Evaluation on Advanced Reactor Machine Learning Subversion Attacks
M3CT-23IN1105013: Autonomous System Inference, Trojan, and Adversarial Reprogramming Attack and Defense

U.S. DEPARTMENT OF ENERGY | Office of NUCLEAR ENERGY

CT-23IN110501: Cyber Security for Nuclear Machine Learning Applications (iv)

Overview

- Cyber-Physical Testbed for Analyzing Autonomous Systems
- ML Based Digital Twins in Autonomous Systems
- Predictive Modeling of I/O in Autonomous System Components
- Simulated Cyber Attacks



Defensive Strategies

- Fortifying against Inference Attacks
- Thwarting Trojan Attacks
- Mitigating Adversarial Attacks
- General Defensive Strategies

M2CT-23IN1105017: Complete Evaluation on Advanced Reactor Machine Learning Subversion Attacks

M3CT-23IN1105013: Autonomous System Inference, Trojan, and Adversarial Reprogramming Attack and Defense

INL/RPT-23-75941 Rev:000

Questions