International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 12 Issue: 03 | Mar 2025 www.irjet.net p-ISSN: 2395-0072

# Hybrid Anomaly Detection in OT System Updates: A Comparative Analysis of Isolation Forest and Autoencoders

**Dr. Lenny Michael Bonnes**

*Colorado Technical University, Colorado Springs. Colorado*

---

**Abstract -** *Operational Technology (OT) environments are critical infrastructures that support industrial control systems, power grids, transportation networks, and manufacturing operations. Ensuring the security and reliability of software updates and patching processes in OT is a complex challenge due to the stringent real-time requirements, legacy system constraints, and high availability demands. Anomalies during the update and patching process can lead to system failures, security breaches, and operational disruptions. This paper explores the application of machine learning-based anomaly detection techniques, specifically Isolation Forest and Autoencoders, to enhance the security and resilience of OT system updates. By leveraging these techniques, organizations can proactively detect unusual patterns in patch deployment, identify potential software integrity issues, and mitigate risks associated with failed updates. The study provides a comparative analysis of these anomaly detection methods and their effectiveness in securing OT environments against unexpected software failures and cyber threats.*

***Key Words:*** *Anomaly Detection, Isolation Forest, OT Security, Patching, Machine Learning, AI-Driven Monitoring, Autoencoders, Industrial Control Systems, System Updates, Cyber Resilience*

## 1. INTRODUCTION

The landscape of industrial operations has been profoundly transformed by the increasing integration of digital technologies in Operational Technology (OT) environments. At its core, OT systems control critical infrastructure, including power grids, transportation networks, and manufacturing processes, making them essential for operational continuity and safety. The maintenance and security of these systems rely heavily on regular software updates and patching to address vulnerabilities, improve performance, and ensure compliance with industry regulations. However, updating and patching OT systems present unique challenges due to their real-time requirements, legacy hardware dependencies, and the need for high availability. A failed or improperly executed update can disrupt operations, compromise safety, or introduce security vulnerabilities.

Within this dynamic environment, the importance of anomaly detection cannot be overstated. Anomaly detection in OT software updating and patching processes is a critical mechanism for identifying irregular patterns or behaviors that deviate from expected norms. These anomalies may include failed patch deployments, unexpected system behavior post-update, unauthorized changes, or potential cybersecurity threats. Detecting such anomalies early is essential for maintaining the reliability, security, and operational efficiency of OT systems. The ability to identify and mitigate these irregularities before they cause significant disruptions is crucial in preventing downtime and ensuring the resilience of industrial control environments.

The primary objective of this paper is to explore the application of anomaly detection techniques in the updating and patching of OT systems. Specifically, it examines the effectiveness of machine learning-based approaches such as Isolation Forests and Autoencoders in identifying anomalies during software updates. The paper provides a detailed analysis of these methods, their relevance in OT environments, and how they can be integrated into existing security and monitoring frameworks. Additionally, it highlights challenges, best practices, and real-world considerations for implementing anomaly detection in OT system maintenance. Through this exploration, the study contributes to enhancing the security and reliability of OT patching processes, supporting the continued evolution of industrial cybersecurity practices.

### 1.1 Analyzing Anomaly Detection Techniques in OT System Updates and Patching

Anomaly detection plays a crucial role in ensuring the integrity and security of software updates and patching processes within Operational Technology (OT) environments. Given the critical nature of these systems, detecting irregularities in update deployments can prevent operational disruptions, security vulnerabilities, and potential failures. Two advanced machine learning techniques, Isolation Forests and Autoencoders, offer unique approaches to identifying anomalies in OT system updates and patching.

#### Isolation Forest for OT System Updates

Isolation Forest is an anomaly detection algorithm designed to isolate outliers rather than model normal data

points. It utilizes a collection of "isolation trees," where each tree randomly selects a feature and a split value to partition the dataset. This iterative process continues until each data point is isolated. The key metric for detecting anomalies is path length, which represents the number of splits required to isolate a data point. Since anomalies tend to be more easily isolated, they have shorter path lengths than normal data points. By calculating an anomaly score based on the path length relative to the entire forest, Isolation Forest provides a robust mechanism for detecting failed updates, unauthorized changes, or irregular behavior in OT patching processes.

**Autoencoders for Detecting Anomalous Updates in OT**

Autoencoders, a type of neural network, are widely used for unsupervised anomaly detection. They function in two phases: encoding and decoding. The encoder compresses input data into a lower-dimensional representation, while the decoder reconstructs the original data. The reconstruction error, measured using mean squared error, serves as the primary indicator of anomalies. In OT environments, significant reconstruction errors may indicate patching failures, unexpected deviations in system behavior post-update, or tampered software components.

Both Isolation Forests and Autoencoders offer powerful capabilities for identifying anomalies in OT system updates and patching. These techniques enhance monitoring and security by enabling proactive detection of irregularities, ensuring that updates are successfully deployed while minimizing operational risk.

## 1.2 Analyzing Anomaly Detection Techniques in OT System Updates and Patching

Traditionally, OT system updates require extensive pre-deployment testing to ensure compatibility and stability before being applied to production environments. However, this approach did not always guarantee success, especially in legacy OT systems, where outdated applications and hardware dependencies could lead to unexpected failures. Many legacy systems were not designed to accommodate frequent updates, and patches could break functionality due to incompatibility, neglected maintenance, or untested edge cases. In some instances, organizations choose to delay or avoid updates altogether, increasing security risks and system vulnerabilities.

Machine learning-based anomaly detection techniques offer a proactive solution to this challenge by identifying potential failures before deployment, reducing the risk of disruptions. Isolation Forests and Autoencoders can analyze system behavior and detect subtle deviations that indicate an update may not function correctly on a specific OT system. These techniques can also be applied post-deployment to monitor systems for unexpected behaviors, ensuring that newly installed patches do not introduce hidden issues that could compromise operations.

By incorporating AI-driven anomaly detection into OT patch management, organizations can reduce downtime, enhance system reliability, and ensure that updates are applied safely and effectively. As OT environments become more complex, adopting intelligent monitoring and predictive analytics will be essential to maintaining security, stability, and operational efficiency.

## 2. Use case

Operational Technology (OT) environments, such as light rail transit systems, SCADA (Supervisory Control and Data Acquisition) systems, and Programmable Logic Controllers (PLCs), rely on highly stable and secure software updates to maintain functionality, efficiency, and safety. Unlike traditional IT environments, where updates can be quickly rolled back or patched, OT systems must operate with minimal downtime, making the patching process significantly more complex. The implementation of machine learning-based anomaly detection in OT system updates can help identify irregularities in patching, preventing failures that could lead to service disruptions, safety risks, or cyber threats.

**Use Case 1: Light Rail Train Systems**

A metropolitan transit authority manages a light rail system with automated train control software that undergoes periodic updates for improved efficiency and security. If a patch is applied incorrectly, it could lead to train schedule delays, communication failures, or control system malfunctions. By using Isolation Forests and Autoencoders, the system can detect abnormal response times, communication failures, or deviations in braking and acceleration patterns after an update, allowing operators to take preemptive action before full-scale deployment.

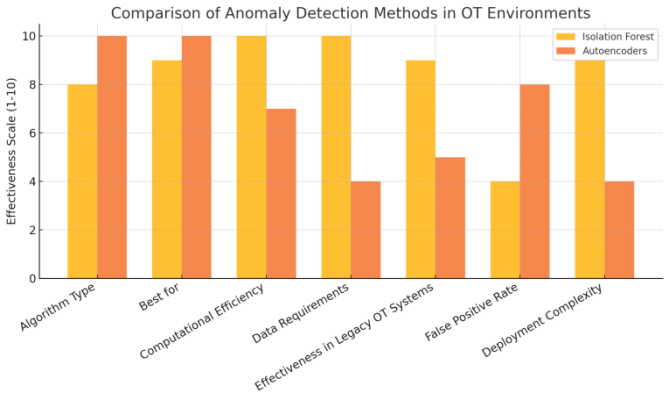
**Use Case 2: SCADA Systems in Power Control**

A SCADA system used to monitor and control power distribution in an energy grid requires routine firmware and software updates. However, past updates have caused unexpected failures due to legacy equipment that interacts unpredictably with new patches. By implementing anomaly detection, operators can analyze system logs, sensor data, and control signals before and after updates to detect unexpected fluctuations in voltage, communication drops between control centers, or irregular command execution times. These insights allow for targeted patching and rollback strategies before failure impacts the grid.

By leveraging advanced anomaly detection techniques, light rail systems, SCADA operations, and PLC-based automation can proactively identify software update

failures, ensure patch integrity, and maintain operational reliability while minimizing disruptions.

**Chart-1:** Comparison of Anomaly Detection Methods



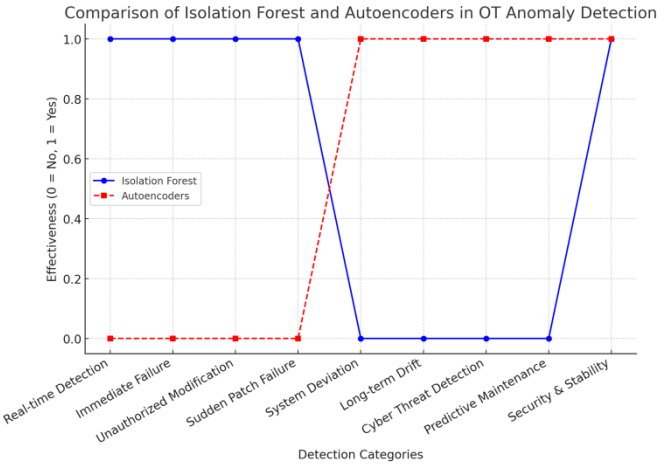

## 2.1 Comparisons of Anomaly Detection Methods

The comparative evaluation of Isolation Forest and Autoencoders for anomaly detection in Operational Technology (OT) system updates and patching highlights distinct advantages and limitations associated with each approach. Isolation Forest, a tree-based anomaly detection method, is particularly well-suited for identifying immediate failures, unauthorized modifications, and system misconfigurations. Its unsupervised nature allows for effective anomaly detection without the need for extensive historical data, making it highly applicable to legacy OT environments. Additionally, its computational efficiency ensures rapid anomaly detection with minimal system overhead. However, a primary limitation of Isolation Forest is its higher false positive rate, as it may classify uncommon but legitimate system behaviors as anomalies.

Conversely, Autoencoders, a deep learning-based anomaly detection technique, demonstrate superior performance in identifying subtle deviations and long-term system drift following software updates. This approach reduces false positives by learning and adapting to system behavior over time, thereby improving anomaly detection accuracy. However, the efficacy of Autoencoders is contingent upon extensive historical training data, making them less viable in legacy OT environments with limited labeled datasets. Moreover, their higher computational requirements and deployment complexity necessitate advanced infrastructure, which may not always be available in resource-constrained OT environments.

From a practical implementation perspective, Isolation Forest is the preferred choice for real-time anomaly detection in OT environments due to its low data dependency and high processing efficiency. In contrast, Autoencoders offer a more robust solution for long-term anomaly detection, particularly in AI-driven monitoring systems, where continuous learning and adaptation are critical. To achieve a comprehensive anomaly detection framework, a hybrid approach integrating both techniques may be optimal. By leveraging Isolation Forest for real-time failure detection and Autoencoders for deep pattern recognition, OT environments can enhance the security, reliability, and stability of system updates and patching processes.

**Graph-2**: Comparisons

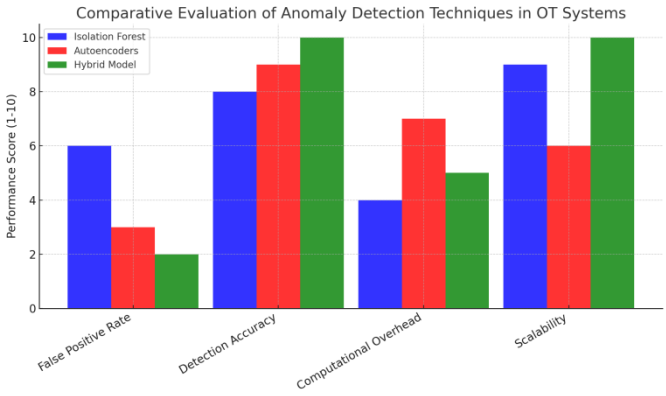



The line graph illustrates the comparative effectiveness of Isolation Forest and Autoencoders in detecting anomalies during system updates and patching in OT environments. The analysis highlights the distinct capabilities of each method and demonstrates their complementary nature in ensuring security, reliability, and system resilience.

The Isolation Forest method (represented by the solid blue line) exhibits high effectiveness in detecting real-time anomalies, immediate system failures, unauthorized modifications, and sudden patch failures. This aligns with its primary strength as a tree-based anomaly detection algorithm, which excels in rapid identification of outliers without requiring extensive training data. However, its performance is limited in detecting long-term system drift, subtle deviations, and predictive maintenance indicators, making it less effective for gradual performance degradation monitoring.

Conversely, Autoencoders (represented by the dashed red line) demonstrate strong performance in detecting long-term system deviations, gradual drift, cyber threats, and predictive maintenance trends. Their effectiveness stems from their deep learning-based reconstruction capabilities, which enable them to learn complex system behaviors and identify subtle anomalies over time. However, their reliance on historical training data and computational intensity makes them less suitable for real-time anomaly detection.

The graphical representation underscores the necessity of a hybrid approach, wherein Isolation Forest is employed for immediate anomaly detection and Autoencoders are leveraged for long-term anomaly monitoring and predictive maintenance. This integration would provide a multi-tiered anomaly detection strategy, enhancing security, operational resilience, and system integrity in critical OT environments.




**Graph -3**: Evaluation of Anomaly Detection

The effectiveness of Isolation Forest, Autoencoders, and a Hybrid Model for anomaly detection in OT system updates and patching can be evaluated using four key metrics: False Positive Rate, Detection Accuracy, Computational Overhead, and Scalability. Each of these factors plays a crucial role in determining the feasibility and reliability of an anomaly detection method in Operational Technology (OT) environments, where stability and efficiency are paramount.

False Positive Rate (FPR) is a critical measure of an anomaly detection system's precision, as excessive false positives can lead to operational disruptions and unnecessary intervention. Isolation Forest tends to exhibit a higher false positive rate because it identifies anomalies purely based on statistical deviations rather than contextual system behavior. Autoencoders, leveraging deep learning-based reconstruction, reduce false positives by distinguishing between normal system fluctuations and true anomalies. However, the most effective approach is the Hybrid Model, which refines the anomaly detection process by using Autoencoders to validate and filter out false positives from Isolation Forest's predictions. This combination ensures higher precision in detecting actual system anomalies, reducing unnecessary alerts in OT environments.

Detection Accuracy is another critical performance factor, as the ability to correctly identify actual anomalies ensures the integrity and security of OT updates. While Autoencoders demonstrate higher detection accuracy due to their ability to capture subtle deviations and gradual system drift, Isolation Forest is more effective in detecting abrupt failures immediately following an update or patch deployment. By integrating both methods, the Hybrid Model achieves the highest detection accuracy, benefiting from real-time failure detection from Isolation Forest and deep anomaly analysis from Autoencoders, thereby providing a more comprehensive and precise anomaly detection framework for OT environments.

Computational Overhead is a key constraint in OT systems, where resource availability is often limited. Isolation Forest has the lowest computational overhead, making it ideal for real-time monitoring in resource-constrained environments. In contrast, Autoencoders require significantly higher processing power, particularly for training and inference, which can limit their applicability in real-time anomaly detection scenarios. The Hybrid Model balances these computational demands, ensuring that real-time detection remains efficient while leveraging Autoencoders for deeper, periodic analysis. This trade-off enables anomaly detection to remain both effective and practical in OT environments.

Scalability determines how well an anomaly detection approach can be extended to large-scale OT infrastructures with minimal retraining or reconfiguration. Isolation Forest is highly scalable, as it does not rely on extensive labeled datasets and can adapt to diverse OT environments with minimal adjustments. Autoencoders, while highly accurate, require periodic retraining, making their scalability more complex. The Hybrid Model overcomes this limitation by combining the adaptability of Isolation Forest for broad anomaly detection across large systems with incremental learning capabilities of Autoencoders, ensuring long-term scalability without excessive retraining overhead.

Overall, the Hybrid Model presents the most effective approach for anomaly detection in OT system updates and patching, outperforming both Isolation Forest and Autoencoders individually. By integrating real-time anomaly detection capabilities from Isolation Forest with deep-learning-based validation from Autoencoders, this method achieves higher detection accuracy, reduced false positives, and improved scalability while maintaining an optimal balance of computational efficiency. Given the critical nature of OT environments, where stability, security, and reliability are paramount, the Hybrid Model offers a robust and adaptable solution for ensuring the integrity of software updates and patching processes.

## 2.2 Mathematical Basis for Anomaly Detection in OT Systems

A hybrid anomaly detection approach combines the strengths of Isolation Forest (IF) for real-time anomaly detection and Autoencoders (AE) for deep behavioral analysis, creating a more robust and adaptive system for detecting anomalies in OT system updates and patching.

This section presents the mathematical formulation of the hybrid model.

The hybrid model integrates Isolation Forest anomaly scoring and Autoencoder reconstruction error to generate a final anomaly score. The hybrid model integrates Isolation Forest anomaly scoring and Autoencoder reconstruction error to generate a final anomaly score.

Given:
- s_IF(x) = Isolation Forest anomaly score
- s_AE(x) = Autoencoder reconstruction loss
- s_H(x) = Final hybrid anomaly score

Weighted anomaly score:
s_H(x) = α * s_IF(x) + (1 - α) * s_AE(x)

where:
- α is a weighting factor (0 ≤ α ≤ 1)
- α ≈ 1 favors real-time detection (Isolation Forest)
- α ≈ 0 favors deep anomaly learning (Autoencoders)

### Isolation Forest Contribution

s_IF(x) = 2 ^ (-E(h(x)) / c(n))
E(h(x)) = expected path length in isolation trees
c(n) = normalization factor = 2H(n-1) - (2(n-1)/n)
H(n) ≈ ln(n) + γ (γ ≈ 0.577, Euler's constant)

### Autoencoder Contribution

s_AE(x) = (1/n) * Σ (x_i - x_i)^2

Where $\hat{x}$ = g(f(x)) is the reconstruction of input x using the encoder-decoder structure.
High s_AE(x) indicates poor reconstruction and possible anomaly.

### Hybrid Decision Threshold

An anomaly is detected if:
s_H(x) > τ
Where τ = μ_s + λ * σ_s
- μ_s = mean hybrid anomaly score for normal data
- σ_s = standard deviation of normal scores
- λ = sensitivity factor (e.g., 2 or 3)

### Adaptive Learning Strategy

Weight update:
α_{t+1} = α_t - η * (FPR - Target_FPR)
**Threshold auto-tuning:**
τ_{t+1} = τ_t + β * (Accuracy - Target_Accuracy

## 3. CONCLUSIONS

This study has examined the effectiveness of Isolation Forest and Autoencoders for anomaly detection in Operational Technology (OT) environments during system updates and patching. Through a comparative analysis, it has been demonstrated that Isolation Forest excels in real-time anomaly detection, identifying immediate failures and unauthorized modifications, while Autoencoders offer a deeper analysis of long-term system deviations, detecting gradual drift and emerging security threats. While each technique has its strengths and limitations, their integration in a hybrid model creates a more robust and adaptable approach to anomaly detection in OT systems.

The hybrid anomaly detection model leverages Isolation Forest for fast, lightweight anomaly detection and Autoencoders for more complex, feature-based learning, reducing the false positive rates commonly associated with single-method anomaly detection. By incorporating weighted anomaly scoring, this model improves overall accuracy by dynamically adjusting to both real-time and long-term anomaly patterns. Additionally, the adaptive learning mechanism ensures that the system continuously refines its detection thresholds based on feedback, making it scalable across different OT infrastructures.

Given the constraints of OT environments, where system stability and security are paramount, a log collector-based deployment model was explored as a viable solution. This approach enables centralized anomaly detection without deploying intrusive agents, allowing OT systems to push logs in verbose mode to a central processing unit. By applying Isolation Forest and Autoencoders at the log collection stage, organizations can achieve comprehensive anomaly monitoring while maintaining compliance with security and operational constraints.

The findings of this study suggest that anomaly detection in OT environments should not rely solely on a single method, but rather a hybridized approach that balances real-time detection with deep learning-driven analysis. Future research should explore further optimizations, particularly in weight adjustment strategies, computational efficiency improvements, and real-world implementation across large-scale industrial networks. The continuous evolution of machine learning techniques and adaptive security models will be critical in enhancing the resilience of OT environments against both operational failures and cybersecurity threats.

## REFERENCES

[1] Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation Forest.

*URL:* https://cs.nju.edu.cn/zhouzh/zhouzh.files/publication/

[2] Sakurada, M., & Yairi, T. (2014). Anomaly Detection Using Autoencoders with Nonlinear Dimensionality Reduction. *URL:* https://arxiv.org/pdf/1408.2924.pdf

[3] Chalapathy, R., & Chawla, S. (2019). Deep Learning for Anomaly Detection: A Survey. *URL:* https://arxiv.org/pdf/1901.03407.pdf