ESP-IJACT
ESP International Journal of Advancements in Computational Technology
ISSN: 2583-8628 / Volume 2 Issue 4 November 2024 / Page No: 62-67
Paper Id: IJACT-V2I4P109 / Doi: 10.56472/25838628/IJACT-V2I4P109

Original Article

Real-Time Anomaly Detection for Insider Threat Prevention in Federal Systems

Hariprasad Sivaraman
Independent Researcher, USA.

Received Date: 02 September 2024 Revised Date: 01 October 2024 Accepted Date: 03 November 2024

Abstract: *Despite being key institutions in both national and state security functions, federal agencies handle incredibly massive amounts of sensitive data, making them a high value vector for insider threats. This demonstrates how insider threats often evade traditional security mechanisms that fail to detect malicious activity in real-time and mitigate risk effectively in a timely manner. A real-time insider threat detection model using machine learning for anomaly detection in federal systems therefore is proposed in this paper. Through predictive analytics, behavioral profiling and ongoing monitoring, this model is intended to protect federal systems from the threat of an internal security breach and improve response time.*

Keywords: *Insider Threats, Federal Systems, Anomaly Detection, Machine Learning, Behavioral Monitoring, Cybersecurity.*

I. INTRODUCTION

Over the past several years, insider threats have emerged as one of the biggest security risks facing federal systems. External threats are searching for entry points within a Web Application Firewall (WAF), whereas insider threats use the access they already have. The threats may include data theft, system sabotage or unauthorized access to sensitive information causing serious risks to the national security. Perimeter security approaches are useless when it comes to insider threads; for the simple reason that they can slip through those defenses with no trouble if they are legitimate users. Another concern is that federal systems have their own security requirements and compliance obligations, which also require specialized solutions for anomaly detection with regulatory guidance. In this paper, a machine learning based real-time anomaly detection framework is proposed that solves the problem of insider threat in federal systems by continuously learning from user behavior and doing adjustment on thresholds to detect very slight behavioral anomalies.

II. PROBLEM STATEMENT

Most federal systems do not incorporate sophisticated near real-time methods for discovering insider threats. Conventional methods deal with after-the-fact investigation, prolonging responses and allowing systems to remain vulnerable. Federal mission needs, call for low-latency detection and mitigation solutions that comply with regulatory standards. The proposed framework aims to:

a) Real-time analysis and classification of user behavior.
b) Operate within regulatory requirements for data privacy.
c) Deliver an intelligent and responsive framework that can process heterogeneous big data streams.

III. PROPOSED SOLUTION: REAL-TIME ANOMALY DETECTION FRAMEWORK

A federal systems-centric, multi-layered anomaly detection solution based on ML using layered architecture (LA) with real-time processing layer and ML-less alternative.

A. System Architecture

The system architecture consists of Data Ingestion, Feature Engineering, Model Training and Selection, and Real-Time Detection and Alerting layers.

- Data Ingestion Layer: Gathers user activity logs from various spheres directly, network traffic, application logs, and authentication records. High-throughput data flow via Apache Kafka or Apache Pulsar, full compliance with GPDR and similar data protection regulation with anonymization and encryption of collected raw data.
- Feature Engineering Layer: This layer derived behavioral metrics such as the length of a session, how often sensitive files were accessed and patterns in access. Keystroke dynamics as an example of the derived indicators for representing user interaction and sentiment analysis based on NLP shows anomalies in language used on communication logs.

- Model Training and Selection Layer: The framework incorporates several ML models adapted for anomaly detection:
    - Isolation Forest and One-Class SVM for unsupervised anomaly detection.
    - LSTM-based Autoencoders for detecting sequential anomalies.
    - Graph Neural Networks (GNNs) examine intricate networked user interactions.
    - XGBoost with Random Forest o Ensemble Model helps identify more subtle behavior
    - Ensemble Models such as Random Forest with XGBoost improve detection of nuanced behaviors.
- Real-Time Detection and Alerting Layer: Real-time application of trained models on each new transaction, therefore batch and speed layers for predicting abnormalities in real time. Scalable model deployment supported with docker containers and k8s support, alerting mechanism comprises of secure channels for alerts.

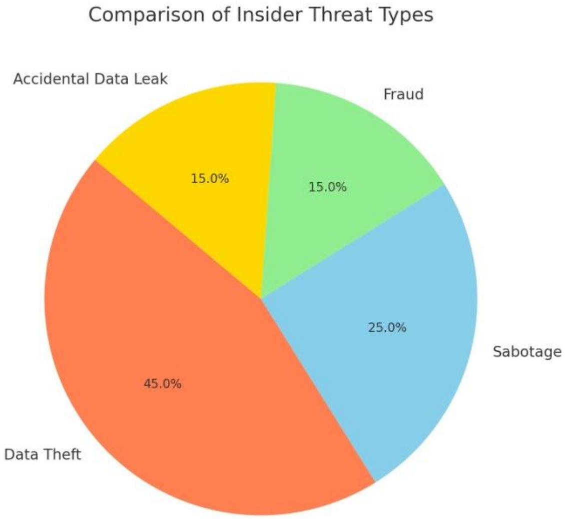

Figure 1: Comparison of Insider Threat Types

B. Behavioral Profiling and Anomaly Detection

- Detects and monitors baseline activity patterns via behavioral profiling, making this the basis of the framework.
- User Clustering: By means of k-means and DBSCAN clusters (to group users): the groups should be identified with their typical behavior. When the behavior deviates significantly from what is normal we flag it as an anomaly.
- Online Learning: The system implements online learning to adjust itself to new data, combined with federated learning for the case where data is available locally without violating privacy.

C. Data Preprocessing and Feature Selection

- Rigorous preprocessing ensures optimal model performance
- Normalization and Encoding helps standardize the inputs while Dimensionality Reduction such as PCA simplifies complex data.
- Algorithms such as RFE for Feature Selection, are helpful to make the processing efficient in real-time.

D. Model Evaluation and Fine-Tuning

Models are rigorously tested for suitability:

- Evaluation Metrics: Preference is given to precision, recall and low number of false negatives.
- Hyperparameter, Optimization: Process such as grid search enhance the model accuracy for online applications.

E. Real-Time Monitoring and Feedback Loop

A feedback loop refines detection accuracy:

- Feedback integration: Alerts that are verified feed back into the model, adjusting its thresholds.
- Explainability: Ensures actions taken by detection tools like SHAP or LIME can be audited through transparent design.

F. Security and Compliance

The framework complies with standards such as FISMA and NIST, ensuring compliance through infrastructure such as encryption and access control.

IV. APPLICATION AND FRAMEWORK

The proposed real-time anomaly detection framework offers federal systems enhanced capabilities to monitor and mitigate insider threats proactively. Below are detailed uses and applications of the framework:

A. Real-Time Threat Detection and Monitoring

Federal agencies can leverage this framework to achieve continuous, real-time threat detection that goes beyond traditional log analysis and manual monitoring. Key features include:

- Activity-Based Anomaly Detection: By continuously checking user behaviors (those deemed of high importance) in real-time, abnormal patterns that may indicate suspicious activities or security events can be detected, such as accessing protected files, executing large data transfers and accessing the system from non-standard locations and devices.
- Dynamic Risk Scoring: The risk score of every user action is assessed through anomaly detection algorithms. Such high-risk activities allow instant alerts to a firm security team for investigation which makes sure that possible threats are detected, and security of data is not compromised before exfiltration or any severe damage is done.
- Incident Response Automation: When integrated with Security Information and Event Management (SIEM) tools, the framework can automate responses such as restricting access, logging off users or prompting two-factor authentication requests when anomalous behavior is identified. The automation part of it simply speeds up response times and contains damage.

B. Enhanced Access Management and Privilege Escalation Monitoring

This framework also strengthens access management and privilege monitoring, two critical areas for mitigating insider threats:

- Privileged User Monitoring: One of the top areas of risk from insider threats is high-level access that privileged users are often provided. The anomaly detection framework keeps an eye on privileged users for any sorts of red flag movements like unauthorized access or large-scale changes to the system components other than regular operational activity. Any departure from historical behavior is flagged and escalated for review.
- Behavior-Based Access Control: By observing patterns in the long term, abstract models to control access are updated that react automatically by increasing or decreasing permissions of certain users. For instance, if a user habitually accesses systems during normal business hours, an attempt to access outside of these hours may prompt additional verification steps.

C. Threat Intelligence and Forensic Analysis

This framework helps in strengthening the threat intelligence database, a resource that is beneficial when defending proactively or handling post-incident forensic investigations:

- Threat Pattern Recognition: The system creates a knowledge base from all the previously flagged incidents, learning new patterns of threat. Having this database allows for faster identification of familiar threats or new variations of insider attacks, thus reinforcing the overall security posture in the long-term.
- Detailed Forensic Trails: The framework provides logs of flagged events, which will help federal agencies perform deep forensics on certain threats. This record consists of time-stamped user actions, access history, and system interaction history that can be extracted to conduct internal audits or regulatory reporting; it can even serve as evidence during any potential legal investigations.

V. IMPACT

Implementing this framework in federal systems can profoundly impact cybersecurity at the national level, enhancing security, operational efficiency, and compliance capabilities.

A. Strengthening National Security

Insider threats are a major concern for federal agencies given that some of the country’s most mission-critical data and infrastructure is handled within their walls. Therefore, this framework with real-time and accurate proactive user behavior monitoring becomes an effective defensive tool against these attacks. From a national security standpoint, this has the following impacts:

- Minimization of Data Breach Risks: The framework minimizes the risk of data breaches that could involve sensitive or classified information if detected early enough in an insider attack, protecting citizen data while also keeping national interests secure.
- Reduction of Disruption Driven by Insiders: Some of the insider threats are intended to disrupt or damage your organizations, not steal data. Ensuring government and national continuity while reducing the chances of insider activity

impacting critical services, public trust, and society itself — early detection of suspicious behavior can only be a good thing.

- Support for Intelligence and Defense Agencies: The framework can be leveraged by intelligence and defense agencies to protect the most highly classified systems, adding another line of defense against breaches of national security data.

B. Cost Savings and Resource Optimization

They can be costly to manage and mitigate some of time, including data breach recovery. The framework provides a range of financial and operational efficiencies which include:

- Reduction of Manual Monitoring Costs: With automated real-time monitoring and alerting, the requirement for large manual monitoring teams is minimized, freeing up security personnel at all levels to focus on higher value work functions such as response and strategic security planning.
- Cut Costs Associated with Incident Response: Analyzing live alerts on anomalous behaviors reduces the risk of expensive incidents (like a data breach). This is also cost-effective, as the recovery after a cyber-attack can be huge for federal agencies.
- Focused Resource Allocation: Only high-risk alerts will catch the attention of resources by using this framework, decreasing alert fatigue and enhancing response teams.

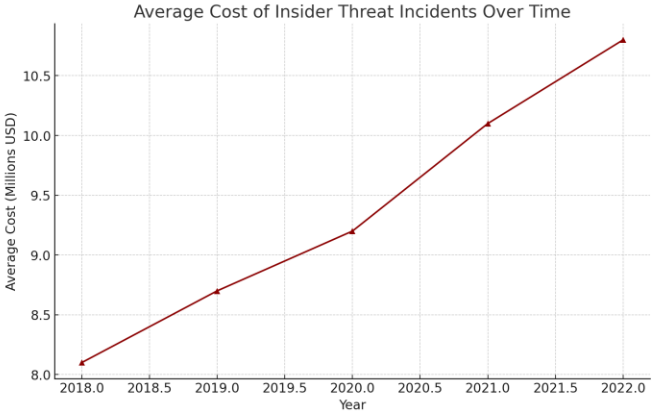

Figure 2: Average Cost of Insider Threat Incidents over Time

C. Enhanced Compliance and Audit Readiness

Federal systems have very different regulatory and compliance mandates that related to things like the Federal Information Security Management Act (FISMA) and National Institute of Standards and Technology (NIST). With a design that's fully poised to implement compliance, the framework ensures:

- Constant Compliance: Data protection standards are designed into the framework, which naturally incorporates elements such as audit trails, encryption and anonymization that support federal compliance. By constantly keeping on with this, the risk related to regulatory penalties and audits is decreased.
- Automation and Audit Trails: Comprehensive logging and audit trail capabilities that allow federal agencies to easily create reports to show they are complying with cybersecurity requirements. Moreover, record-keeping automation provides support for yearly audits and compliance checks by making the auditing process easier with constant transparent, verifiable records of security activities.

VI. SCOPE

The scope of this framework extends beyond federal systems, with potential applications in various sectors and government levels. This section explores the framework's scalability, adaptability, and areas of extension.

A. Scalability and Adaptation Across Federal Agencies

The framework has been designed to be modular and scalable, adapting across federal departments with limited reconfiguration:

- Cross-Agency Play: This framework can be adapted across federal agencies including defense and public health; and is customized for unique security postures and operational workflows. Designed with a flexible architecture that facilitates connection with various systems and datasets, it can easily integrate with an agency's unique cybersecurity needs.

- Pilot Programs and Phased Rollouts: Agencies might consider performing an assessment run before using the framework at high security departments or sensitive areas. Agencies can phase in development of specific problems in order to further improve the framework prior to full adoption.
- Scalable Microservices Architecture: The microservices architecture also allows the framework to scale along with federal agencies as they grow, or as more systems are integrated. This flexibility allows the framework to stay relevant when volumes grow, or an agency builds on its operations.

B. Extension to State and Local Governments

Although this framework is primarily aimed at federal systems, state and local governments contend with the same insider threat challenges and would benefit from these insights. The framework can allow state governments with such confidential data but in two neighboring countries to monitor user access evil and find if there is anyone doing so could be framed. This especially applies to departments that are responsible for social services, public health records or criminal justice information.

Municipalities operating a public utility or local infrastructure (e.g., water) could use a distilled version of the above framework. These sectors are becoming more digital and cyber threats continue to grow, and the framework provides a means of safeguarding vital infrastructure from a local perspective.

C. Adaptation to Private Sector Applications

This flexibility in the framework makes it ideal for private sector organizations managing sensitive data and those who are likely to experience insider threats. Possible uses include:

- Financial Sector Security: The framework can be used by financial institutions to safeguard crucial data such as client records and financial deals being frequent insider threat targets. Behavioral profiling and anomaly detection features of the Image Recognition Framework are especially useful for detecting fraudulent or unauthorized activities.
- Data Protection in Health: The health sector manages large volumes of sensitive patient data. Such a framework may help healthcare providers with fulfilling these requirements where activity can be monitored for users accessing EHRs and if unusual behavior is detected, raises an alert. Techniques that preserve privacy, such as federated learning, ensure patient confidentiality.
- Defense Contractors and Critical Industries: Private sector defense contractors that support federal agencies can leverage this framework to maintain security of proprietary data. This allows these contractors to protect intellectual property while enhancing national security and lowering the risk of black-hat data leaks by a disgruntled insider.

VII. CONCLUSION

It is important to subject federal systems that run sensitive and high-stakes data to real-time anomaly detection for preventing insider threats. Federal agencies rely on traditional security controls that are targeted outward and cannot detect, let alone respond to, insider threats in time to prevent SYS/BI from occurring. Firstly, high risk in terms of subtle patterns; those that tend not to be caught as being suspect, clearly identifying insider threats by utilizing machine learning algorithms and behavioral profiling however more importantly this framework uses a multi-layered approach detector.

The inclusion of models such as Isolation Forest combined with LSTM autoencoders, and Graph Neural Networks provides an end-to-end framework for accurate and real-time detection that not only improves accuracy but also flexibility. In addition, the continual-learning elements built into this framework will allow it to evolve along with new threat patterns and methods of operation, reducing false positives over time and increasing detection accuracy.

Not only does this framework bolster national security but there are substantial monetary and logistical advantages as well. It allows real-time alerts and automated response mechanisms to mitigate the response times therefore, minimizing the costs associated with insider threat incidents. Also, its certification to federal cybersecurity standards helps agencies comply with regulations and keep sensitive information secure.

In short, this scalable and compliance-ready framework for real-time anomaly detection is a proactive system specifically developed for the individual needs of federal systems. In light of the changing nature of insider threats, this approach helps federal agencies to defend more effectively and strengthens resilience for national infrastructure while protecting valuable data against damaging attacks. Further research and implementation may consider the application of this framework in conjunction with cross-agency threat intelligence networks to develop a stronger defense against insider threats.

VIII. REFERENCES

[1] C. Aljawarneh, M. Aldwairi, and M. B. Yassein, “Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model,” *Journal of Computational Science*, vol. 25, pp. 152-160, 2018.

[2] E. Bertino and G. Ghinita, “Towards mechanisms for detection and prevention of data exfiltration by insiders: Key challenges and potential approaches,” in *Proc. IEEE Int. Conf. on Privacy, Security, Risk and Trust (PASSAT)*, 2011, pp. 139-146.

[3] M. Bishop and C. Gates, “Defining insider threat,” in *Proc. IEEE Computer Society Symposium on Security and Privacy Workshops (SPW)*, 2014, pp. 225-232.

[4] M. Umer, A. Sher, H. Jan, K. Ullah, and A. Zaman, “Modeling suspicious insider threat using structural anomalies in social networks,” *Computers & Security*, vol. 94, pp. 101-115, 2020.

[5] N. J. Hulst and T. A. LeClair, “Continuous security monitoring for insider threat detection: A scalable approach for the enterprise,” *Security Journal*, vol. 33, no. 4, pp. 511-524, 2020.

[6] Verizon. "2023 Data Breach Investigations Report (DBIR)." Available at: [https://www.verizon.com/about/news/media-resources/attachment?fid=65e1e3213d633293cd82b8cb].

[7] Ponemon Institute. "2022 Cost of Insider Threats Global Report." Available at: [https://go.proofpoint.com/rs/309-RHV-619/images/Ponemon_2022Report_A4_Final_UK.pdf].

[8] IBM. "2023 Cost of a Data Breach Report." Available at: [https://www.ibm.com/security/data-breach].

[9] Carnegie Mellon University. CERT Insider Threat Center. "Insider Threat Report." Available at: [https://resources.sei.cmu.edu/].

[10] Carnegie Mellon University. CERT Insider Threat Test Dataset. Available at: [https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=508099].

[11] Deloitte. "2022 Insider Threat Report." Available at: [https://www2.deloitte.com/global/en.html].