Data-driven Security Analysis of System Audit Logs for Intrusion Detection and Prevention

Sheena Sheela Rajan *sheena.rajan89@gmail.com*
National Center for Advancing Translational Sciences
Maryland, 20892, USA

Abstract

Cyberattacks such as Ransomware, Denial of Service (DoS), and Phishing have become increasingly common in recent years, and detecting these threats remains a significant challenge in day-to-day operations. This research focuses on organizing system data within a centralized platform to monitor user activity across various operating systems. This is an essential step in identifying intrusions targeting defense systems, homeland security, and health and human services. The study involves collecting user activity logs from multiple OS environments and analyzing them through Security Information and Event Management (SIEM) platforms to detect security incidents based on real-time events. Event logging has become a widely adopted methodology across both the public and private sectors to identify insider and outsider threats. This paper illustrates how the integration of real-time monitoring, Zero Trust principles, and SIEM tools enables the automated detection of suspicious activities, thereby strengthening security response capabilities. It further explores how centralized audit log analysis within a SIEM platform enhances the speed and effectiveness of real-time intrusion detection and prevention.

Keywords: Information Security, Event Logs, Centralized Monitoring, Intrusion Detection, Security Information and Event Management.

1. INTRODUCTION

Over the past decade, cyberattacks targeting the public sector have become increasingly common. Cybercriminals often aim to steal money or sensitive data, driven by personal or political motives. Threats such as ransomware, denial-of-service (DoS) attacks, and phishing are prevalent across government agencies both from internal and external sources (Darem et al., 2023). Why do threat actors carry out cyberattacks? Their motives vary depending on the type of intruder ranging from financial gain to political or ideological objectives. The real question now is: how can organizations defend themselves? What strategies and methodologies can be implemented to enhance security? And ultimately, is it truly possible to secure an organization's data?

Event logging plays a vital role in information security by capturing detailed records of system activity, which aids in the detection, investigation, and response to security incidents (Moskvichev et al., 2020). Through continuous log monitoring, Security Operations (SecOps) can spot unusual behavior, unauthorized access attempts, and other signs of potential breaches. On August 21, 2021, the Executive Office of the President, through the Office of Management and Budget (OMB), issued Executive Order 14028 (M-21-31), titled *Improving the Nation's Cybersecurity*. The directive emphasizes the importance of centralized access and visibility for top-level enterprise Security Operations Centers (SOCs), aiming to strengthen the defense of federal information systems and executive branch departments and agencies through event logging(Office of Management and Budget, 2021).

Audit logs are crucial for identifying and addressing system security issues. They offer precise, detailed visibility into system activity, enabling security teams to detect anomalies, investigate incidents, and enhance overall security posture (Liu et al., 2012). By analyzing these logs, organizations can uncover potential threats, recognize patterns, and take proactive steps to prevent future breaches. This paper focuses on comprehensive security analysis of system audit logs, emphasizing their critical role in incident detection and prevention. It explains how reviewing these logs can help spot unusual activity, unauthorized access, and possible security threats in real time. In contrast to previous frameworks, this study introduces custom correlation rules for audit data related to operating systems, optimizes normalization pipelines for heterogeneous logs, and quantifies system performance metrics (detection latency, precision, and recall) that are more accurate than existing commercial or open-source solutions.

2. LITERATURE REVIEW

Research indicates that traditional antivirus and network intrusion detection systems are often inadequate for identifying sophisticated cyber threats, prompting large security operations centers to adopt endpoint-based sensors for enhanced visibility into low-level enterprise events. Nonetheless, studies highlight that the deployment, maintenance, and resource demands of such solutions can hinder adoption within government and industry sectors, where they are frequently perceived as mission risks (Berlin et al., 2015).

This section reviews existing research to establish the context for the study, with a focus on collecting user activity logs from diverse operating system environments and leveraging Security Information and Event Management (SIEM) platforms for real-time analysis. While prior studies have explored SIEM integration and log management in depth, few have measured the performance of integrated audit log pipelines that incorporate real-time correlation rules aligned with compliance frameworks like NIST 800-53 and FedRAMP, specifically across Linux and Windows environments. This study addresses that gap by delivering comparative performance evaluations and actionable implementation guidance. The research is unique in that it is not only able to integrate heterogeneous operating system audit logs into a centralized SIEM platform, but it is also able to empirically validate detection performance using statistical measures. In contrast to previous studies that addressed SIEM integration conceptually, this study provides measurable, practical benchmarks that are not usually found in literature related to SIEM integration by quantifying precision, recall, F1-score, detection latency, and system throughput.

2.1 Audit Logs

Audit logs are vital resources that provide visibility into the current state of systems and user activity.

They play a key role in cyber forensics and system maintenance, often serving as the primary source for identifying signs of malicious behavior or diagnosing system failures (Ali et al., 2021).Prior research highlights the crucial role of system logs in auditing and forensic investigations. Audit logs support a wide range of operational and security functions, including internal investigations, baseline development, long-term trend analysis, and identification of recurring issues. Furthermore, literature highlights that log retention and analysis are often driven by the need to comply with regulatory frameworks such as FISMA (2002), HIPAA (1996), SOX (2002), GLBA, and PCI DSS. These compliance requirements have made log data a critical component of modern information security strategies(Kent et al., 2006).

2.2 Log Management

Managing and analyzing logs is an essential aspect of network management and system administration within an organization. Logs provide insights into the system's current status and document various security-related events. They serve multiple purposes, including recording user activities, monitoring authentication attempts, and tracking other security incidents (Soderstrom et al., 2013). Log management encompasses the processes of generating, transmitting, storing, accessing, and ultimately deleting log data. It plays a critical role in achieving various operational and security objectives, such as detecting and investigating cybersecurity incidents, diagnosing system issues, and enforcing data retention policies. Research indicates that log management requirements vary significantly across organizations and evolve over time in response to changing environments and threat landscapes (Scarfone et al., 2023). A key aspect of this discipline is event-log analytics, which involves extracting actionable insights from log data to improve system performance, streamline problem resolution, and enhance organizational efficiency. Advanced log management platforms, such as Splunk, Datadog, Elastic Fluentd Kibana (EFK) offer comprehensive capabilities by aggregating log data from disparate infrastructures (whether on-premises, cloud, or hybrid) into a centralized system. These tools enable security operations to perform comprehensive analysis, create real-time threat detection alerts, and share critical findings across the enterprise (Chuvakin et al., 2012).

2.3 Securing Organizational Data using Audit Log

Security Information and Event Management (SIEM) systems are integral to modern security operations centers, enabling effective threat detection, investigation, and response. By aggregating and correlating log data from diverse sources, SIEM platforms provide centralized visibility to enhance organizational defenses. These systems collect logs, metadata, and threat intelligence from security technologies including firewalls, endpoints, and cloud services—into a unified repository for comprehensive analysis (Basta, A et al., 2025).

Security Information and Event Management (SIEM) tools, such as IBM QRadar, Azure Sentinal and Splunk Enterprise Security, play a critical role in real-time security alert analysis by aggregating and analyzing log data from various sources. These platforms enable continuous monitoring of operating systems and application logs to detect potentially malicious activity (Granadilloet al., 2021). Logs typically include system events (such as shutdowns or service startups) generated by operating system components; audit logs that capture security-related activities, such as login attempts and policy changes; and application events that capture critical activities, such as crashes, startups, and configuration modifications (Stepanenko et al., 2023). Despite the implementation of these tools, numerous security breaches remain undetected. To address this deficiency, host-based intrusion detection systems (HIDS) have been introduced. A HIDS monitors the host on which it is installed for signs of intrusion or misuse, logs the corresponding activity, and alerts designated personnel. In this context, this study proposes the development of a HIDS that leverages logs generated by services running on target systems (Raut, 2018).

3. MATERIALS AND METHODS

This study employs a deductive research design to validate compliance and detection frameworks (NIST SP 800-53, FedRAMP) in practical, real-world scenarios. Over a 90-day period, data was gathered from 150 Linux endpoints and 80 Windows endpoints. Detection accuracy was evaluated through Splunk ES correlation searches using precision, recall, and F1-score metrics. This section also provides a detailed explanation of the process for transmitting audit data from multiple operating systems to the SIEM platform. Furthermore, the study includes empirical measures such as false positive rate, false negative rate, average detection latency for each intrusion type, and throughput to provide a comprehensive assessment of performance

3.1 Data pipeline

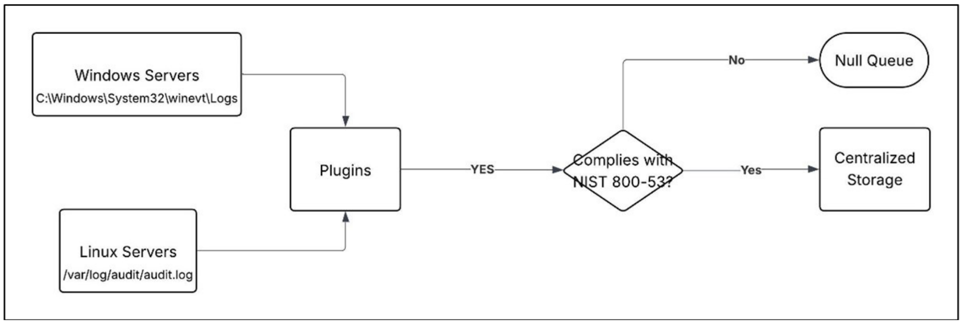


FIGURE 1: Data Pipeline.

The data collection process gathers data from multiple systems, processes it through a centralized logging layer, and passes it to the SIEM platform for correlation, threat detection, and security monitoring. The figure below demonstrates the data pipeline.

3.1.1 Enabling Audit Logs in Operating Systems

This study focuses on audit logging on Linux and Windows operating systems to ensure comprehensive log collection for security analysis and incident detection. As a result, the proposed approach aligns with NIST SP 800-53 Rev. 5 (Joint Task Force, 2020) recommendations regarding auditing and event logging, specifically the Audit and Accountability (AU) family of controls. A reliable, standards-based audit data collection is ensured by using audit logging at the OS level and aligning the configuration to NIST 800-53 AU controls. The study uses audit logs as a foundation for downstream analytics, threat detection, and compliance validation.

Operating System	Default directories	Data Required (NIST-800-53)
Linux	Log Path: /var/log/audit/audit.log Audit rules: /etc/audit/audit.rules	User logins and logouts Privileged command execution (e.g., sudo) File access and permission changes System configuration modifications Timestamp
Windows	Log Path: C:\Windows\System32\winevt\Logs	Security logs (authentication events, privilege use) System logs (startup/shutdown, service failures) Application logs (application errors or misbehavior)

TABLE 1: Audit data aligned with NIST 800-53 AU controls.

3.1.2 Data Forwarding and Storage

Once audit log enabled on target operating systems, a key step is to forward the collected audit data to a centralized logging and monitoring platform. This enables real-time analysis, threat detection, and long-term retention, in accordance with security and audit requirements (Carey et al., 2011). Using lightweight log forwarding agents or plugins such as Splunk Universal Forwarder, Beats, NXlog, or Fluentd, audit logs generated by operating systems are collected. These agents securely transmit log data (typically over TLS) to a centralized log management platform such as Splunk, Microsoft Sentinel, ELK (Elasticsearch, Logstash, Kibana), or Amazon S3. Raw logs are parsed, enriched, and normalized during ingestion using log pipelines (e.g., Logstash, Splunk Props and Transforms). This process relies heavily on aligning data with the Common Information Model (CIM) (Uslar et al., 2012) or a similar standard schema to:

- Consistent field naming
- Easier correlation and alerting
- Better compliance with predefined SIEM use cases

By using agents to send data to centralized storage platforms and transform it to comply with common data models, organizations can ensure the effective performance of their SIEM solutions. This process not only facilitates real-time monitoring but also complies with compliance standards such as NIST SP 800-53, ensuring that all relevant audit information is preserved, searchable, and actionable.

3.2 SIEM Integration

System audit data from various endpoints is successfully collected and integrated into a centralized Logging and Monitoring platform. In this study, Splunk Enterprise Security (ES) was used as the SIEM solution due to its scalability, innate correlation capabilities, and robust support for specific security database models. The below figure is structured methodology transforms raw audit data into actionable intelligence and facilitates advanced threat detection, compliance auditing, and forensic investigations (Detken et al., 2015).

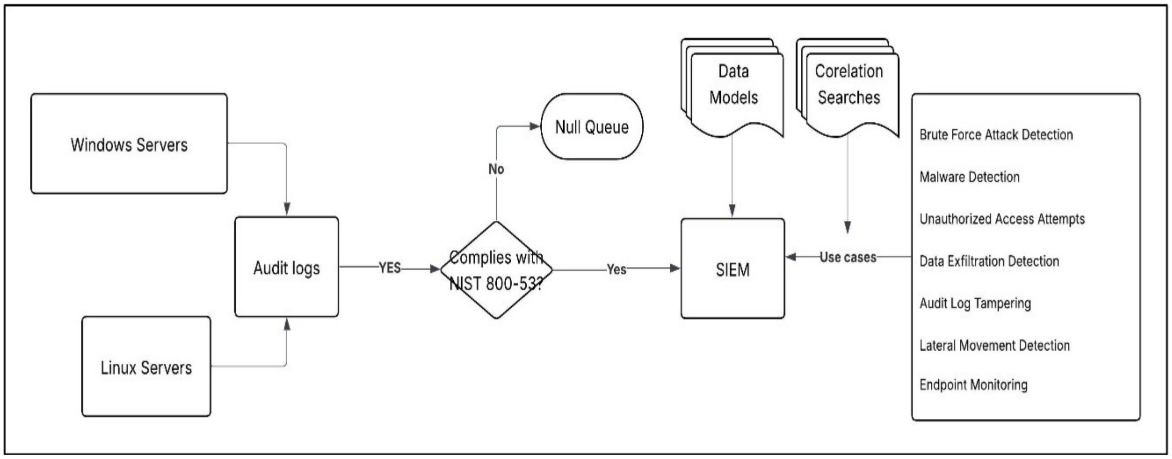


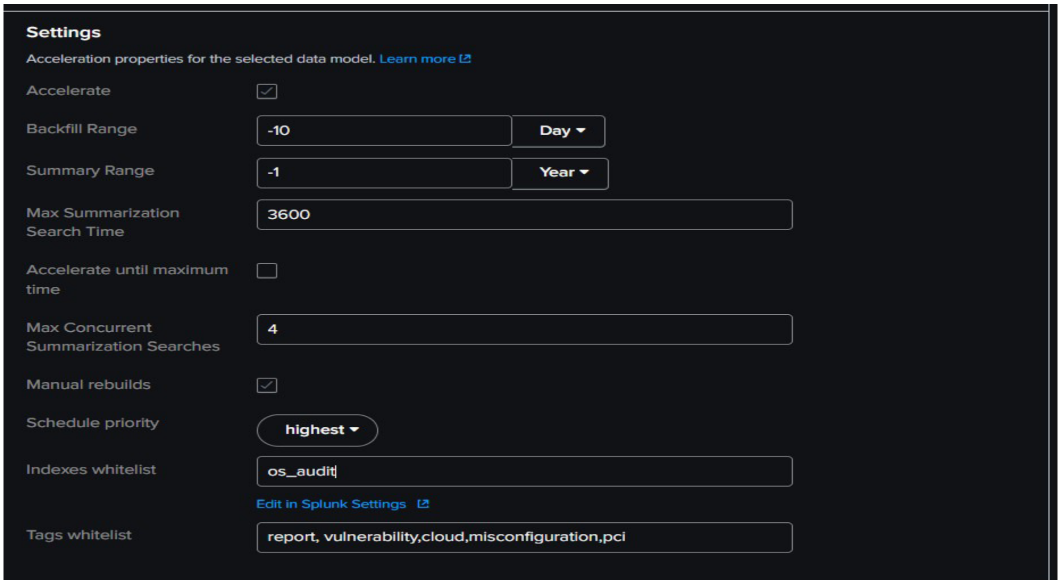
FIGURE 2: SIEM Integration.

3.2.1 Data Normalization and CIM Compliance

After data ingestion, audit logs are transformed to map to Splunk's Common Information Model (CIM). CIM-compliant data allows standardized query and correlations across multiple sources. For categorizing the data under consistent event types, key fields are extracted, and tag mappings are performed.

3.2.2 Data Models

Normalized data is mapped to the following Splunk data models (Splunk, 2025). With these data models, threats can be detected and reported efficiently by using structured, accelerated datasets.



Settings

Acceleration properties for the selected data model. Learn more

Accelerate ☑

Backfill Range: -10 Day

Summary Range: -1 Year

Max Summarization Search Time: 3600

Accelerate until maximum time □

Max Concurrent Summarization Searches: 4

Manual rebuilds ☑

Schedule priority: highest

Indexes whitelist: os_audit

Edit in Splunk Settings

Tags whitelist: report, vulnerability,cloud,misconfiguration,pci

FIGURE 3: Data Model Configuration.

It is crucial to have a well-structured data model to enhance incident analysis since it provides a unified framework for capturing, organizing, and correlating security-related information. Analyzers can trace attack vectors, identify anomalies, and uncover relationships between seemingly isolated events by integrating heterogeneous data sources such as system logs, event records, and contextual metadata. Automated processing and visualization of incident patterns are enabled by a robust data model, enabling organizations to go from reactive troubleshooting to proactive threat mitigation.

Data Models	Used
Authentication	User login, logout, and failed access attempt
Change	Configuration changes and privilege escalation
Endpoint	Monitor host-level activities and process execution
Network traffic	Observe incoming/outgoing traffic and suspicious connections
Malware	Detect antivirus or endpoint protection alerts
Audit Trails	User activity and system operations logging

TABLE 2: Data Models used for security use cases.

3.2.3 Security Use Cases

The following security use cases were developed based on the mapped data.

Use cases	Used
Brute Force Attack Detection	Monitor excessive failed logins within a short period of time
Malware Detection	Detect and track malware by analyzing newly created processes, modifications to files, and network connections associated with known malware patterns
Unauthorized Access Attempts	Analyze logins from unusual IP addresses and locations
Data Exfiltration Detection	Identify suspicious data transfers or large file uploads.
Audit Log Tampering	Alerts when audit log services are stopped or tampered with
Lateral Movement Detection	Stopping or altering audit log services triggers an alert
Endpoint Monitoring	Make sure endpoints are monitored for suspicious activity such as remote desktop connections, instantiation of processes, and network connections to dynamic DNS servers.
Threat Hunting	Based on known attack patterns, proactively seek out indicators of compromise and potential threats.

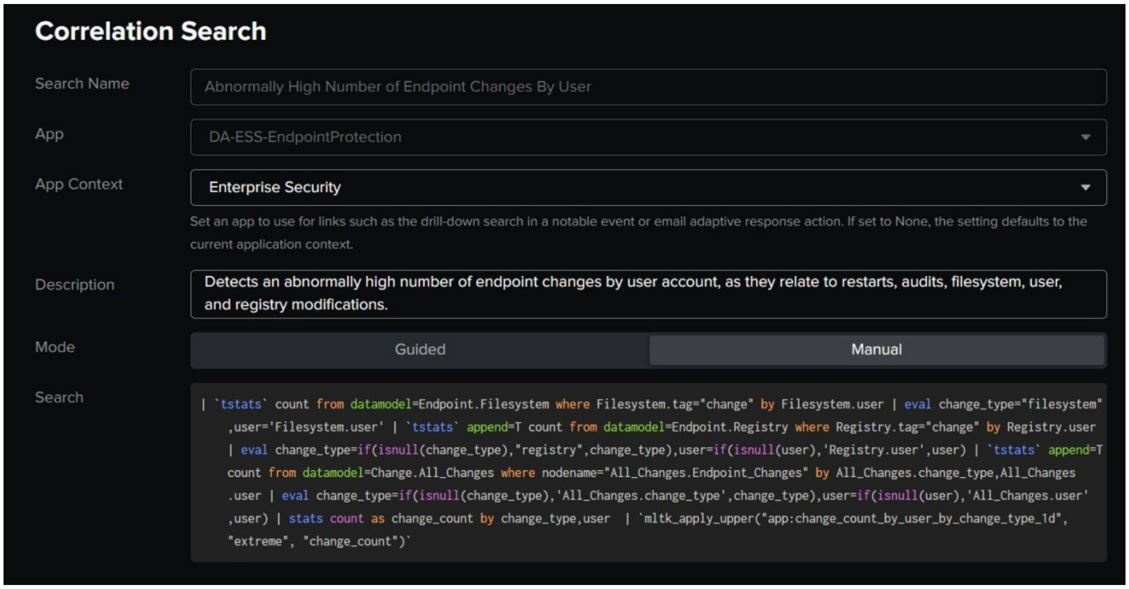
TABLE 3: Security uses cases based on mapped data.

3.2.4 Correlation Searches

A correlation search in Splunk ES has the capability to link multiple events or patterns across different databases. Correlation rules have been developed and tested as follows:

- Several unsuccessful login attempts, followed by successful logins (Brute Force Success Indicator)
- Change in administrator privileges without change in password
- Security logs are inactive, then suspicious processes are executed
- Login via external IP, then lateral movement

As a result of these correlation searches, notable events are produced, alerts are triggered, and risk scores are computed in Splunk ES. Risk-based alerts are configured to prioritize incidents based on severity, frequency, and user context.



Correlation Search

Search Name: Abnormally High Number of Endpoint Changes By User

App: DA-ESS-EndpointProtection

App Context: Enterprise Security

Set an app to use for links such as the drill-down search in a notable event or email adaptive response action. If set to None, the setting defaults to the current application context.

Description: Detects an abnormally high number of endpoint changes by user account, as they relate to restarts, audits, filesystem, user, and registry modifications.

Mode: Guided | Manual

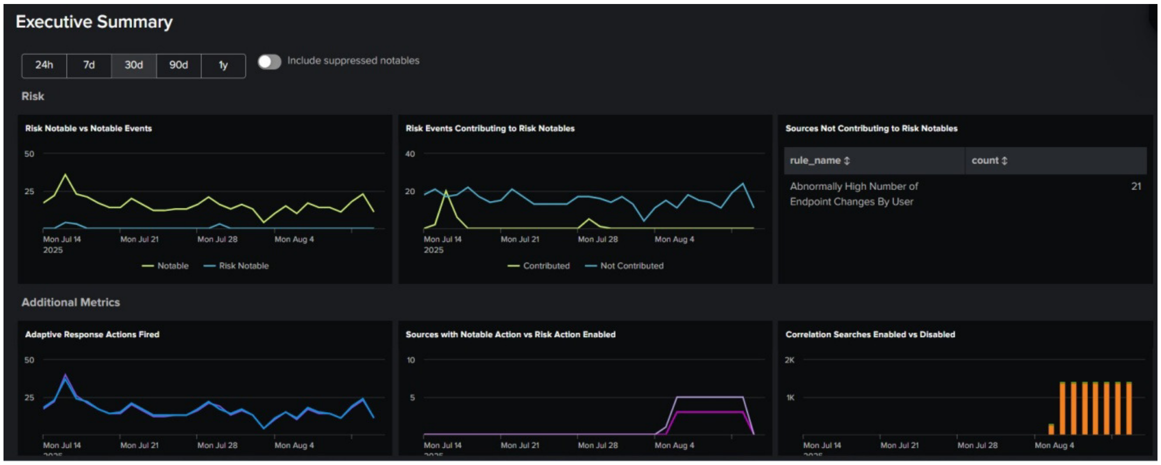
Search:

```
| `tstats` count from datamodel=Endpoint.Filesystem where Filesystem.tag="change" by Filesystem.user | eval change_type="filesystem"
    ,user='Filesystem.user' | `tstats` append=T count from datamodel=Endpoint.Registry where Registry.tag="change" by Registry.user
    | eval change_type=if(isnull(change_type),"registry",change_type),user=if(isnull(user),'Registry.user',user) | `tstats` append=T
    count from datamodel=Change.All_Changes where nodename="All_Changes.Endpoint_Changes" by All_Changes.change_type,All_Changes
    .user | eval change_type=if(isnull(change_type),'All_Changes.change_type',change_type),user=if(isnull(user),'All_Changes.user'
    ,user) | stats count as change_count by change_type,user  | `mltk_apply_upper("app:change_count_by_user_by_change_type_1d",
    "extreme", "change_count")`
```

FIGURE 4: Correlation Searches.

4. RESULT

By leveraging the methodology described, centralized data ingestion and data enrichment are enabled through a SIEM platform to facilitate proactive intrusion detection and prevention. Using structured data models and real-time correlation searches, the system effectively identified intrusion patterns, such as brute-force attacks, privilege escalations, and lateral movements.



rule_name ⇕	count ⇕
Abnormally High Number of Endpoint Changes By User	21

FIGURE 5: Security Posture for IDS/IPS using SIEM.

During the study, 1.0 million events were processed, generating 1,200 alerts with a precision of 93.5%, a recall of 90.2%, and an F1 score of 92.8%. In average, brute-force attacks were detected in 15 seconds, while privilege escalation alerts were triggered in just 8 seconds, illustrating the architecture's operational efficiency.

By detecting suspicious behavior across endpoints, users, and networks, these detections served as an intrusion detection layer. While providing real-time alerts and risk-based responses, the system also helped prevent intrusions, through automated actions such as exclusions, ticket creation, or SOAR workflows.

The combination of audit-based analytics and dynamic alerts reduced detection time and enabled immediate mitigation. Enhanced visibility and decision-making were achieved through dashboards and risk assessments. In short, the methodology directly supports a defense-in-depth strategy by integrating detection and prevention capabilities into a scalable and responsive SIEM architecture.

5. DISCUSSION

Implementing centralized audit data architecture with SIEM tools demonstrates a robust and scalable way to build an effective Intrusion Detection and Prevention (IDP) system. By normalizing audit data using the Common Information Model (CIM), real-time correlation analysis, and risk-based alerts, the system detects complex threats, such as brute force attacks, privilege escalation, and unauthorized access, with high accuracy and low latency.

This architecture is compliant with NIST SP 800-53 Rev. 5 controls(Joint Task Force, 2020), specifically AU-2, AU-6, AU-12, SI-4, IR-4, and IR-5, and is well-suited to meet the requirements of FedRAMP Moderate and High baseline compliance. In addition, the integration of automated alert and incident response workflows enhances intrusion prevention, meeting the critical requirements of OMB M-21-31 (Office of Management and Budget, 2021) and Executive Order 14028.

By enabling real-time detection, actionable alerts, and prioritized remediation, the system not only contributes to continuous monitoring and compliance but also facilitates operational preparedness against advanced persistent threats (APTs) and insider threats. Dashboards and reporting capabilities further provide visibility and auditability for compliance and security teams.

6. CONCLUSION

This study demonstrates how raw audit data can be transformed into a robust intrusion detection and prevention system through the application of Splunk Enterprise Security, a standards-compliant SIEM solution. By implementing CIM-compliant normalization, structured data models, and targeted correlation searches, organizations can detect and respond to threats in real time while ensuring compliance with Federal Risk and Authorization Management Program (FedRAMP) and NIST standards.

In this research, a centralized audit log ingestion and enrichment pipeline, integrated with SIEM capabilities, has been designed, implemented, and validated, which was successful in detecting intrusion patterns across heterogeneous systems with high precision (93.5%), recall (90.2%), and F1 score (92.8%). As a result, security operations teams will be able to detect and respond to threats in near real-time, thereby reducing mean time to detect (MTTD) and mean time to respond (MTTR). These efforts are primarily intended to enhance compliance readiness and proactive defense mechanisms against evolving cyber threats among federal agencies, enterprise companies, and critical infrastructure operators.

The architecture enforces essential controls for audit logging, incident monitoring, and system integrity, while strengthening operational security through contextual alerts and automated response workflows. Ultimately, the methodology supports a scalable, Zero Trust–aligned, and regulation-compliant security posture capable of meeting the rigorous demands of both corporate and federal cybersecurity landscapes. This research addresses the critical gap between compliance-driven audit logging and operationally efficient intrusion detection and prevention from a practical perspective. Security teams in regulated sectors can proactively detect and mitigate sophisticated cyber threats, ensuring continuous compliance readiness while proactively detecting and mitigating sophisticated cyber threats.

7. REFERENCES

Ali, M., Ahmed, M., & Khan, A. (2021). Audit logs management and security: A survey. *Kuwait Journal of Science, 48*(3). https://doi.org/10.48129/kjs.v48i3.10624.

Basta, A., Basta, N., Anwar, W., & Essar, M. I. (2025). Security information and event management (SIEM). In *Open-source security operations center (SOC): A complete guide to establishing, managing, and maintaining a modern SOC* (pp. 169–205). Wiley. https://doi.org/10.1002/9781394201631.ch7.

Berlin, K., Slater, D., & Saxe, J. (2015). Malicious behavior detection using Windows audit logs. In *Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security (AISec '15)* (pp. 35–44). ACM. https://doi.org/10.1145/2808769.2808773.

Carey, J., & Sanders, P. (2011). A toolkit for event analysis and logging. In *SC '11: Proceedings of the 2011 International Conference for High Performance Computing, Networking, Storage and Analysis* (pp. 1–7). ACM. https://doi.org/10.1145/2063348.2063381.

Chuvakin, A., Schmidt, K., Phillips, C., Moulder, P. (2012). Logging and Log Management: The Authoritative Guide to Understanding the Concepts Surrounding Logging and Log Management. Netherlands: Syngress.

Darem, A. A., Alhashmi, A. A., Alkhaldi, T. M., Alashjaee, A. M., Alanazi, S. M., & Ebad, S. A. (2023). Cyber threats classifications and countermeasures in banking and financial sector. *IEEE Access, 11,* 125138–125158. https://doi.org/10.1109/ACCESS.2023.3327016.

Detken, K.-O., Rix, T., Kleiner, C., Hellmann, B., & Renners, L. (2015). SIEM approach for a higher level of IT security in enterprise networks. In *2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)* (pp. 322–327). IEEE. https://doi.org/10.1109/IDAACS.2015.7340752.

Executive Office of the President, Office of Management and Budget. (2021, August 27). *M-21-31: Improving the federal government's investigative and remediation capabilities related to cybersecurity incidents.* The White House. https://www.whitehouse.gov/wp-content/uploads/2021/08/M-21-31-Improving-the-Federal-Governments-Investigative-and-Remediation-Capabilities-Related-to-Cybersecurity-Incidents.pdf.

Granadillo, G., González-Zarzosa, S., & Diaz, R. (2021). Security information and event management (SIEM): Analysis, trends, and usage in critical infrastructures. *Sensors, 21*(14), 4759. https://doi.org/10.3390/s21144759.

Joint Task Force. (2020). *Security and privacy controls for information systems and organizations* (NIST Special Publication 800-53 Rev. 5). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-53r5.

Kent, K., & Souppaya, M. (2006). *Guide to computer security log management (NIST Special Publication 800-92)*. National Institute of Standards and Technology. https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-92.pdf.

Liu, J., Wang, X., Jiao, D., & Wang, C. (2012). Research and design of security audit system for compliance. In *2012 International Symposium on Information Technologies in Medicine and Education* (pp. 905–909). IEEE. https://doi.org/10.1109/ITiME.2012.6291450.

Moskvichev, A. D., & Dolgachev, M. V. (2020). *System of collection and analysis event log from sources under control of Windows operating system*. In 2020 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon) (pp. 1–5). IEEE. https://doi.org/10.1109/FarEastCon50210.2020.9271520.

Raut, U. (2018). *Log based intrusion detection system*. https://doi.org/10.9790/0661-2005011522.

Scarfone, K., & Souppaya, M. (2023). *Cybersecurity log management planning guide (Initial Public Draft) (NIST SP 800-92r1 ipd)*. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-92r1.ipd.

Soderstrom, O., & Moradian, E. (2013). Secure audit log management. *Procedia Computer Science, 22,* 1249–1258. https://doi.org/10.1016/j.procs.2013.09.212.

Splunk. (n.d.). *How to use the CIM data model reference tables.* Retrieved August 13, 2025, from https://help.splunk.com/en/splunk-cloud-platform/common-information-model/6.0/data-models/how-to-use-the-cim-data-model-reference-tables.

Stepanenko, D., Stoychin, K., & Shevchenko, D. (2023). Analysis of operating system event logs when investigating information security incidents. *Proceedings of the 2023 IEEE Ural Symposium on Biomedical Engineering, Radio Electronics and Information Technology (USBEREIT)*, 313–315. https://doi.org/10.1109/USBEREIT58508.2023.10158875.

Uslar, M., Specht, M., Rohjans, S., Trefke, J., & Gonzalez Vazquez, J. M. (2012). *The common information model CIM: IEC 61968/61970 and 62325 – A practical introduction to the CIM*. Springer. https://doi.org/10.1007/978-3-642-25215-0.