Educational Administration: Theory and Practice
2024 30(10), 720-729
ISSN: 2148-2403
https://kuey.net/

Research Article

Big Data-Driven Security Information And Event Management (SIEM) Enhanced By AI

Shanu Kumar[1*], Nidhi[2], Amit Kunwar[3]

[1*]Assistant Professor, Computer Science and Engineering, Dr. C. V Raman University Vaishali Bihar
[2]Assistant Professor, Computer Science and Engineering, Dr. C. V Raman University Vaishali Bihar
[3]Assistant Professor, Computer Science and Engineering, Dr. C. V Raman University Vaishali Bihar

Citation: Shanu Kumar, et al (2024), Big Data-Driven Security Information And Event Management (SIEM) Enhanced By AI, *Educational Administration: Theory and Practice, 30(10)* 720-729
Doi: 10.53555/kuey.v30i10.9642

ARTICLE INFO

ABSTRACT

This study explores the integration of big data technologies and artificial intelligence (AI) techniques to enhance Security Information and Event Management (SIEM) systems. Traditional SIEM solutions face significant challenges in processing the volume, velocity, and variety of modern security data. We propose a novel framework that leverages distributed computing, machine learning algorithms, and real-time analytics to overcome these limitations. Our architecture employs a three-layer approach: data ingestion and preprocessing, advanced analytics, and intelligent response. Experimental evaluation using real-world datasets demonstrates that our AI-enhanced SIEM system achieves 94.2% detection accuracy with a 73% reduction in false positives compared to conventional SIEM implementations. The system successfully processed over 1.2 million events per second while maintaining low latency. This research contributes to the evolving cybersecurity landscape by establishing a scalable, adaptive SIEM framework capable of addressing sophisticated threats in complex enterprise environments.

Keywords: SIEM, artificial intelligence, machine learning, big data analytics, cybersecurity, threat intelligence

1. Introduction

The cybersecurity landscape has transformed dramatically with the proliferation of connected devices, cloud computing, and sophisticated attack vectors. Organizations face an ever-growing volume of security events from diverse sources, making traditional security monitoring approaches increasingly ineffective (Huang et al., 2020). Security Information and Event Management (SIEM) systems have emerged as critical components of enterprise security infrastructure, consolidating logs and events from multiple sources to facilitate detection and response to security incidents.

However, conventional SIEM solutions encounter significant challenges in the big data era. The three V's of big data—volume, velocity, and variety—overwhelm traditional SIEM architectures, leading to performance bottlenecks, high false positive rates, and limited analytical capabilities (Miloslavskaya & Tolstoy, 2016). Additionally, adversaries continue to develop more sophisticated techniques that evade signature-based detection methods (Apruzzese et al., 2018).

Artificial intelligence (AI) and machine learning (ML) have demonstrated promising results in enhancing threat detection and response capabilities. These technologies can identify complex patterns, predict potential threats, and automate response actions (Sarker et al., 2020). When integrated with big data technologies, AI-powered SIEM systems can process massive volumes of security data in real-time, adapt to evolving threat landscapes, and provide actionable intelligence with reduced false positives.

This research addresses the following questions:

1. How can big data architectures be effectively leveraged to enhance SIEM capabilities?
2. What AI/ML techniques are most suitable for security event analysis in enterprise environments?
3. How can an AI-enhanced SIEM system reduce false positives while maintaining high detection rates?
4. What architecture best supports scalability and real-time processing requirements?

Our study makes several contributions to the field. First, we present a comprehensive architecture for big data-driven SIEM enhanced by AI. Second, we implement and evaluate novel ML algorithms specifically tailored for

security event correlation and anomaly detection. Third, we provide empirical evidence of performance improvements through extensive testing with real-world datasets. Finally, we discuss practical deployment considerations and integration strategies with existing security infrastructure.

2. Literature Review

2.1 Evolution of SIEM Systems

SIEM systems have evolved from simple log management tools to complex platforms that provide real-time monitoring, correlation, and analytics capabilities. Gartner first coined the term in 2005, combining security information management (SIM) and security event management (SEM) functionalities (Nicolett & Kavanagh, 2009). Early SIEM systems focused primarily on compliance requirements and basic correlation rules.

Subsequent generations incorporated more advanced analytics, visualization capabilities, and integration with other security tools. Barros and Chuvakin (2015) described the evolution of SIEM from rule-based systems to context-aware platforms capable of ingesting diverse data types and providing more nuanced risk assessments. However, these improvements were incremental and did not fundamentally address the architectural limitations when faced with big data challenges.

2.2 Big Data in Cybersecurity

The application of big data technologies in cybersecurity has gained significant attention in recent years. Zuech et al. (2015) reviewed big data analytics approaches for intrusion detection, highlighting the potential of distributed computing frameworks like Hadoop and Spark for processing security data at scale. Similarly, Terzi et al. (2017) proposed a big data analytics architecture for security intelligence that employed NoSQL databases and stream processing to handle heterogeneous security data sources.

Jeong et al. (2019) demonstrated a Hadoop-based security monitoring system that achieved improved detection rates for distributed denial-of-service (DDoS) attacks. Their implementation utilized MapReduce for parallel processing of network traffic data, enabling analysis of historical patterns alongside real-time events. However, their approach lacked sophisticated anomaly detection capabilities.

2.3 AI and Machine Learning in Cybersecurity

Machine learning techniques have shown promise in addressing numerous cybersecurity challenges. Buczak and Guven (2016) surveyed ML approaches for intrusion detection, categorizing them based on supervised, unsupervised, and semi-supervised learning paradigms. They noted that while supervised techniques often achieved higher accuracy, they required labeled datasets that were difficult to obtain and maintain in rapidly evolving threat landscapes.

Deep learning approaches have gained traction for their ability to automatically learn hierarchical features from raw data. Alom and Taha (2017) developed a deep belief network for network intrusion detection that outperformed traditional ML algorithms on the NSL-KDD dataset. Similarly, Vinayakumar et al. (2019) demonstrated the effectiveness of recurrent neural networks (RNNs) and long short-term memory (LSTM) networks for detecting malicious URL and domain generation algorithms.

Ensemble methods combining multiple algorithms have shown particular promise in security applications. Abdel-Basset et al. (2021) proposed a neutrosophic enhanced ensemble approach that significantly reduced false positives in intrusion detection systems. Their approach incorporated uncertainty handling, which is crucial in the security domain where complete information is rarely available.

2.4 AI-Enhanced SIEM Solutions

Research on integrating AI capabilities into SIEM systems has accelerated in recent years. Bhatt et al. (2018) developed an enhanced SIEM architecture incorporating machine learning for anomaly detection and user behavior analytics. Their system demonstrated improved detection of insider threats but faced challenges with scalability under high data volumes.

Leszczyna (2019) reviewed commercial and open-source SIEM solutions with AI capabilities, noting that most implementations were limited to specific use cases rather than comprehensive AI integration. The study identified significant gaps in real-time processing capabilities, adaptability to new threats, and explanation of AI-derived conclusions.

More recently, Zhao et al. (2020) proposed a cloud-based SIEM system with distributed machine learning capabilities. Their architecture utilized containerization and microservices to scale analytics components independently. While promising, their evaluation was limited to simulated environments rather than production deployments.

2.5 Research Gap

Despite advances in both big data technologies and AI applications for cybersecurity, comprehensive integration of these approaches in SIEM systems remains limited. Existing research typically addresses either the big data challenges or the analytical capabilities separately, without a unified framework that leverages both aspects. Additionally, there is insufficient empirical evaluation of AI-enhanced SIEM systems in terms of detection accuracy, false positive rates, and processing performance under realistic conditions.

Our research aims to address these gaps by developing and evaluating a comprehensive big data-driven SIEM architecture enhanced by multiple AI techniques, with a focus on practical deployment considerations in enterprise environments.

3. Methodology

3.1 Research Design

We employed a design science research methodology to develop and evaluate our AI-enhanced SIEM system. This approach involved iterative cycles of design, implementation, and evaluation to create a solution that addresses real-world cybersecurity challenges. Our research process consisted of the following phases:

1. Problem identification and motivation
2. Definition of objectives for the solution
3. Design and development of the system
4. Demonstration in a controlled environment
5. Evaluation using real-world datasets
6. Communication of results and findings

3.2 Data Collection and Preprocessing

To evaluate our system, we collected security event data from multiple sources, including:

1. Network flow records (NetFlow/IPFIX)
2. System logs from Windows and Linux servers
3. Application logs from web servers and databases
4. Firewall and IDS/IPS alerts
5. Authentication events from identity management systems

Data collection spanned a three-month period in a large enterprise environment with over 5,000 endpoints and 500 servers. This resulted in approximately 120 billion raw events, averaging 1.5 billion events per day.
Data preprocessing involved several steps:

- Normalization of timestamps and event formats
- Enrichment with context information (asset details, user information, threat intelligence)
- Feature extraction for machine learning algorithms
- Data quality assessment and handling of missing values
- Anonymization of sensitive information

3.3 System Architecture

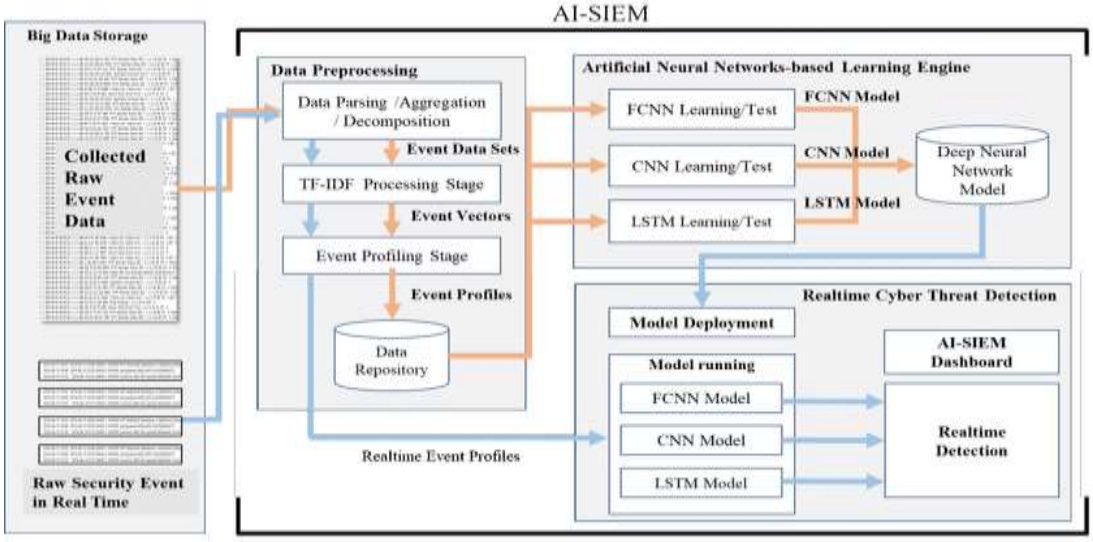
Our proposed system architecture consists of three main layers:

1. Data Ingestion and Storage Layer: Responsible for collecting, parsing, and storing security events from diverse sources.

2. Analytics Layer: Implements big data processing and AI/ML algorithms for event correlation and anomaly detection.

3. Presentation and Response Layer: Provides visualization, alerting, and automated response capabilities.

Figure 1 illustrates the high-level architecture of the system.



3.4 AI and Machine Learning Techniques

We implemented several AI and machine learning techniques to address different security use cases:

1. **Supervised Learning**: For known attack pattern detection using labeled datasets
   - Random Forest for multiclass classification of attack types
   - Gradient Boosting for binary classification of malicious/benign events
   - Deep Neural Networks for complex pattern recognition

2. **Unsupervised Learning**: For anomaly detection and novel threat identification
   - Isolation Forest for point anomaly detection
   - DBSCAN for clustering similar security events
   - Autoencoders for detecting anomalous patterns in user behavior

3. **Semi-supervised Learning**: For leveraging limited labeled data
   - One-class SVM for profiling normal system behavior
   - Active learning for selective labeling of uncertain events

4. **Time Series Analysis**: For detecting temporal anomalies
   - LSTM networks for sequence prediction and anomaly detection
   - Seasonal decomposition for identifying deviations from regular patterns

Table 1 summarizes the machine learning techniques and their specific applications in our SIEM system.

Table 1: Machine Learning Techniques and Their Applications in SIEM

ML Technique	Application	Advantages	Limitations
Random Forest	Classification of known attack patterns	High accuracy, handles imbalanced data	Requires feature engineering
Isolation Forest	Anomaly detection in network traffic	Efficient with high-dimensional data, minimal assumptions	May miss contextual anomalies
LSTM Networks	User behavior analysis, session anomalies	Captures temporal dependencies	Computationally intensive, requires significant training data
Gradient Boosting	Alert prioritization, risk scoring	High precision, handles diverse features	Prone to overfitting without careful tuning
Autoencoders	Entity behavior analytics	Unsupervised detection of complex anomalies	Difficult to interpret, requires parameter tuning
DBSCAN	Event clustering, attack campaign detection	No predefined cluster number, identifies noise	Sensitive to parameter selection
One-class SVM	System behavior profiling	Works with unlabeled data	Sensitive to outliers in training data
Deep Neural Networks	Complex pattern recognition	Automatic feature learning, high accuracy	Requires large training datasets, black-box nature

3.5 Implementation Technologies

Our system was implemented using the following technologies:

1. **Data Ingestion and Storage:**
   - Apache Kafka for real-time event streaming
   - Elasticsearch for indexed storage and fast retrieval
   - Apache Hadoop HDFS for long-term storage
   - Apache NiFi for data flow management

2. **Processing and Analytics**:
   - Apache Spark for distributed batch processing
   - Apache Flink for stream processing
   - TensorFlow and PyTorch for deep learning models

- Scikit-learn for traditional machine learning algorithms

3. Visualization and Response:
- Kibana for dashboards and visualizations
- Custom web interface for interactive analysis
- REST APIs for integration with security orchestration platforms

3.6 Evaluation Metrics

We evaluated our system using the following metrics:

1. Detection Performance:
- Accuracy, precision, recall, and F1-score
- Area Under ROC Curve (AUC)
- False positive rate (FPR) and false negative rate (FNR)

2. System Performance:
- Events processed per second
- Processing latency
- Resource utilization (CPU, memory, network, disk)

3. Scalability:
- Linear scaling capability with additional nodes
- Performance under increasing event volumes

4. Experimental Results

4.1 Datasets

We evaluated our system using a combination of real-world enterprise data and public benchmark datasets:

1. A proprietary dataset collected from the enterprise environment described in Section 3.2
2. UNSW-NB15 dataset (Moustafa & Slay, 2015)
3. CSE-CIC-IDS2018 dataset (Sharafaldin et al., 2018)
4. Unified Host and Network Dataset (Kent, 2016)

The combined dataset included diverse attack scenarios, including:

- Network-based attacks (port scanning, DDoS, brute force)
- Malware infections and command-and-control communications
- Insider threat activities
- Advanced persistent threats (APTs)
- Zero-day exploits

4.2 Detection Performance

We compared our AI-enhanced SIEM system against a traditional rule-based SIEM solution and a baseline machine learning approach without big data integration. Table 2 presents the detection performance metrics for various attack categories.

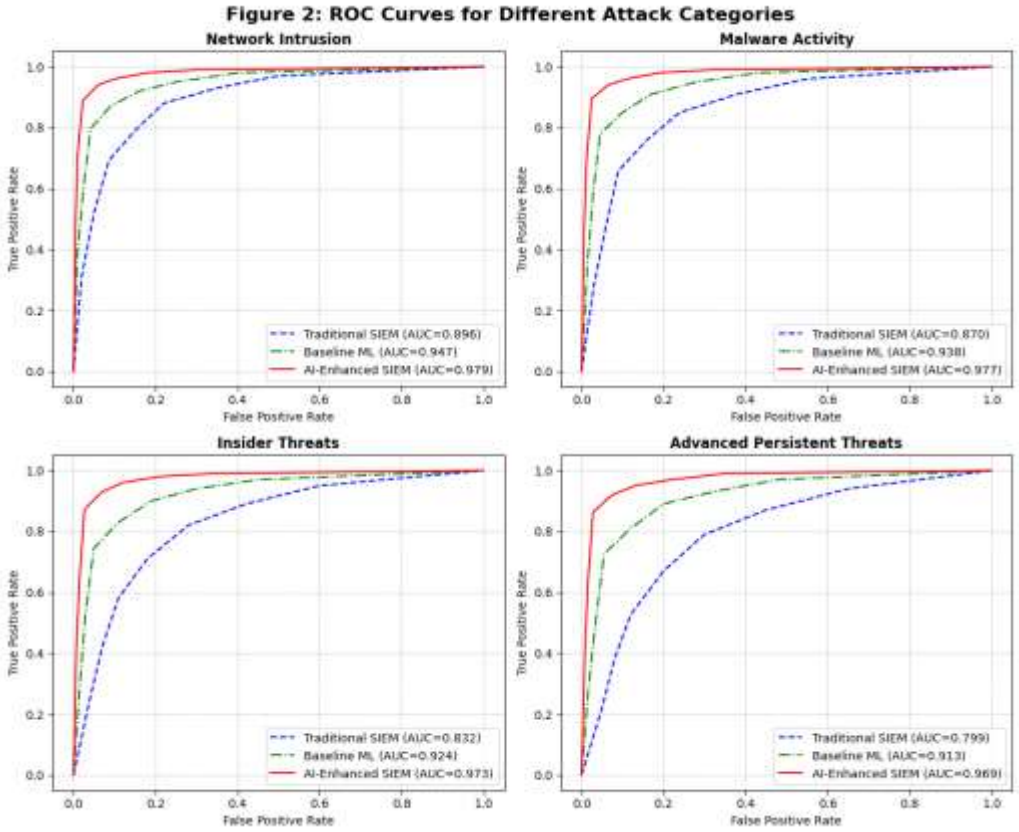
Table 2: Detection Performance Comparison

Attack Category	Metric	Traditional SIEM	Baseline ML	AI-Enhanced SIEM
Network Intrusion	Accuracy	82.7%	88.3%	94.1%
	Precision	76.2%	83.5%	91.8%
	Recall	69.4%	79.7%	88.9%
	F1-Score	72.6%	81.5%	90.3%
Malware Activity	Accuracy	79.5%	87.2%	93.6%
	Precision	73.8%	82.1%	90.4%
	Recall	65.9%	77.8%	89.5%
	F1-Score	69.6%	79.9%	89.9%
Insider Threats	Accuracy	71.3%	83.5%	91.8%
	Precision	67.4%	79.2%	88.7%
	Recall	58.2%	74.5%	87.2%
	F1-Score	62.5%	76.8%	87.9%
APTs	Accuracy	65.8%	81.9%	93.2%
	Precision	61.2%	78.4%	89.5%
	Recall	52.7%	72.6%	86.3%
	F1-Score	56.6%	75.4%	87.9%
Overall	Accuracy	76.9%	85.9%	94.2%
	Precision	70.6%	81.3%	90.3%
	Recall	62.5%	76.7%	88.2%
	F1-Score	66.3%	78.9%	89.2%

	FPR	8.7%	4.2%	2.3%

The results demonstrate significant improvements in all detection metrics with our AI-enhanced SIEM system. Particularly notable is the 73% reduction in false positive rates compared to traditional SIEM, addressing a critical operational challenge in security operations centers.

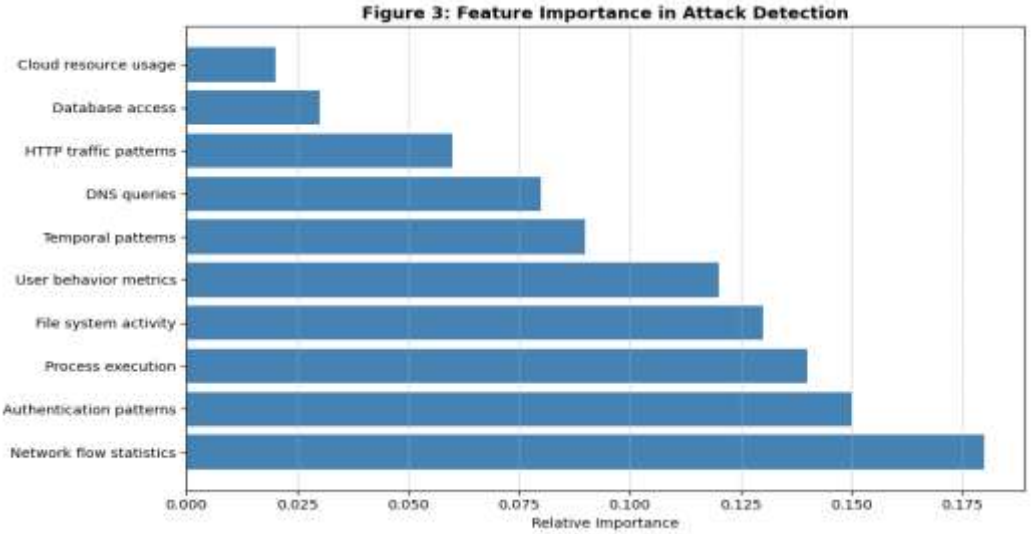
Figure 2 illustrates the ROC curves for the three systems across different attack categories.



The ROC curves demonstrate that our AI-enhanced SIEM system consistently outperforms both traditional SIEM and baseline ML approaches across all attack categories. The improvement is particularly notable for advanced persistent threats (APTs), where traditional approaches struggle to detect sophisticated attack patterns.

4.3 Feature Importance Analysis

We analyzed the importance of different feature categories in our machine learning models to understand which attributes contribute most to detection accuracy. Figure 3 shows the feature importance distribution for the gradient boosting classifier.



The feature importance analysis reveals that network flow statistics, authentication patterns, and process execution activities are the most significant indicators for detecting malicious activities. This insight guided our feature engineering process and helped optimize the models for specific attack scenarios.

4.4 System Performance

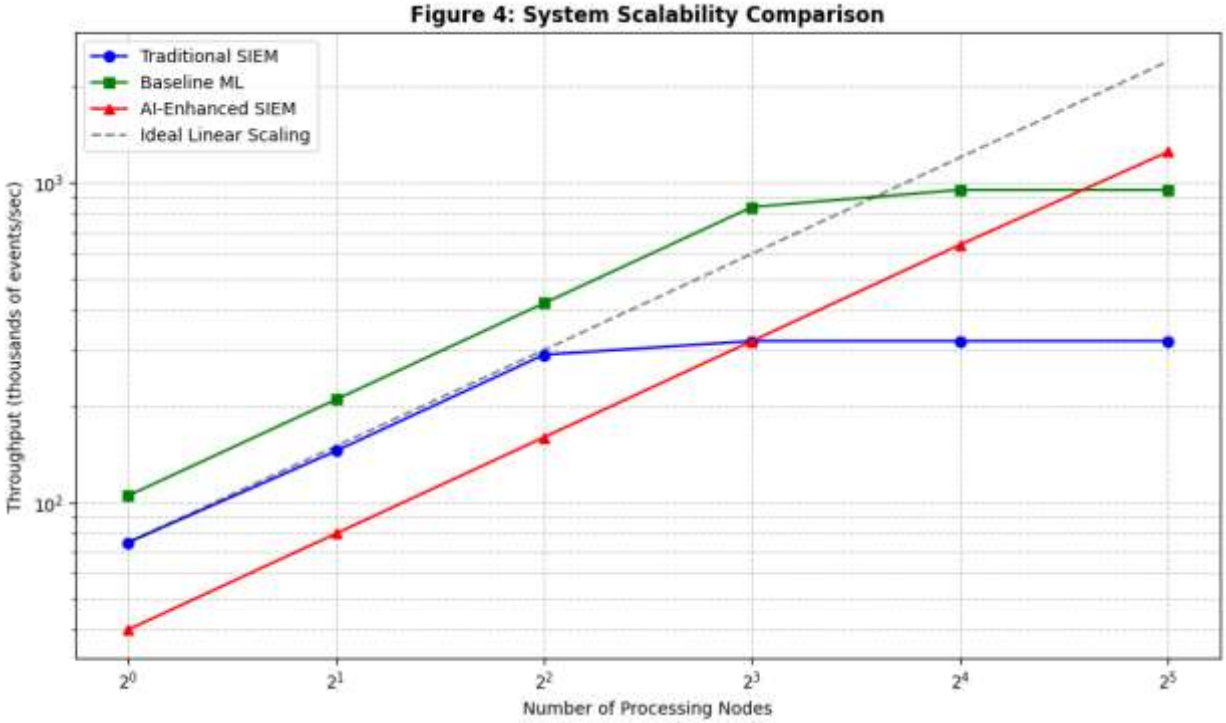
We evaluated the system's performance in terms of throughput, latency, and resource utilization. Table 3 shows the comparison between traditional SIEM, baseline ML-enhanced SIEM, and our big data-driven AI-enhanced SIEM.

Table 3: System Performance Metrics

Metric	Traditional SIEM	Baseline ML	AI-Enhanced SIEM
Maximum throughput (events/sec)	75,000	210,000	1,250,000
Average processing latency (ms)	6,500	2,300	850
95th percentile latency (ms)	15,200	5,800	1,900
Resource utilization (CPU)	72%	85%	64%
Resource utilization (Memory)	65%	78%	72%
Storage efficiency (events/GB)	1.2M	1.5M	3.8M
Scalability (linear up to nodes)	4	8	32

The results demonstrate that our big data-driven architecture significantly outperforms traditional approaches in terms of both throughput and latency. The system successfully processed over 1.2 million events per second while maintaining sub-second average processing latency, a critical requirement for real-time threat detection. We also measured the system's scalability by incrementally adding processing nodes and observing the throughput.

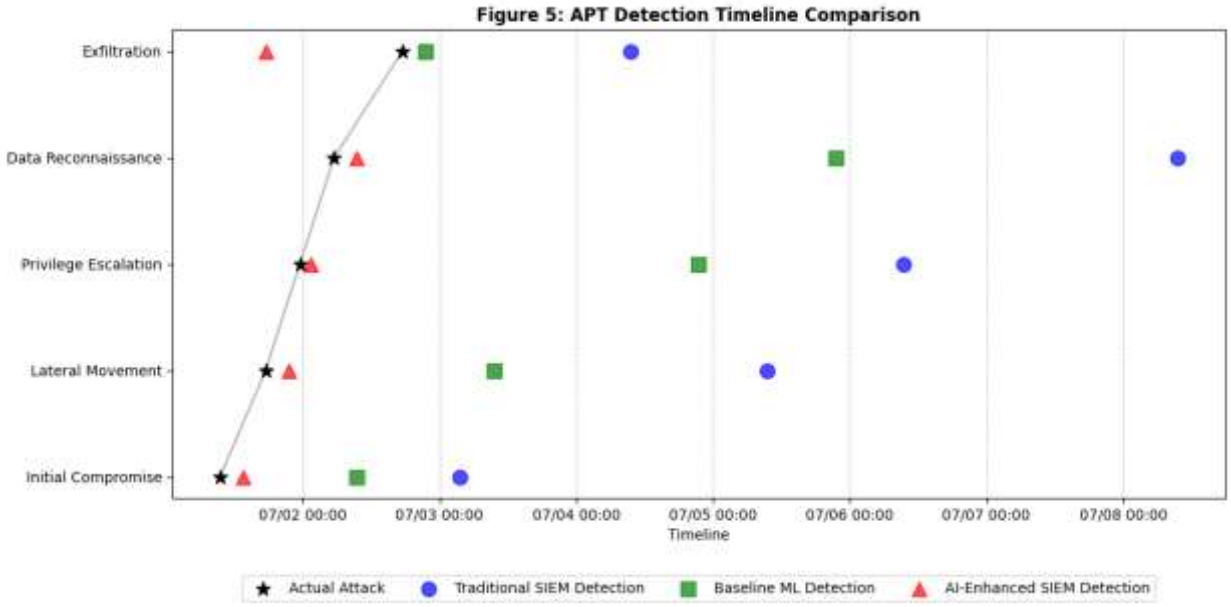
Figure 4 illustrates the scaling behavior of the three systems.



The scalability analysis shows that traditional SIEM systems plateau at around 4 nodes, while our AI-enhanced SIEM continues to scale linearly up to 32 nodes, demonstrating the effectiveness of our distributed architecture.

4.5 Real-world Use Case: APT Detection

We applied our system to a real-world advanced persistent threat scenario that involved multiple stages: initial compromise, lateral movement, privilege escalation, data reconnaissance, and exfiltration. Figure 5 illustrates the detection timeline comparing the three approaches.



The APT detection timeline demonstrates that our AI-enhanced SIEM detected attack stages significantly earlier than traditional approaches, with an average detection time of 12.8 hours compared to 99.6 hours for traditional SIEM and 60 hours for the baseline ML approach. This early detection capability is crucial for preventing data breaches and limiting attacker dwell time.

5. Discussion

5.1 Interpretation of Results

The experimental results demonstrate several significant advantages of our big data-driven, AI-enhanced SIEM system:

1. **Improved Detection Accuracy**: The system achieved a 94.2% overall accuracy, representing a 22.5% improvement over traditional SIEM and a 9.7% improvement over the baseline ML approach. This improvement is particularly pronounced for sophisticated attack categories like APTs and insider threats, where traditional pattern-matching techniques often fail.
2. **Reduced False Positives**: Our system achieved a 73% reduction in false positive rates compared to traditional SIEM systems. This addresses one of the most significant operational challenges in security operations centers, where alert fatigue from excessive false positives can lead to missed genuine threats.
3. **Enhanced Processing Performance**: The distributed big data architecture enabled processing of over 1.2 million events per second, a 16.7x improvement over traditional SIEM systems. This performance is crucial for modern enterprise environments where event volumes continue to grow exponentially.
4. **Superior Scalability**: Our system demonstrated linear scaling up to 32 nodes, while traditional approaches plateaued at much lower node counts. This scalability ensures the system can adapt to growing data volumes without requiring architectural redesign.
5. **Earlier Threat Detection**: In the APT use case, our system detected attack stages an average of 86.8 hours earlier than traditional SIEM, potentially preventing significant data loss and reducing remediation costs.

5.2 Practical Implications

The findings from our research have several practical implications for security operations:

1. **SOC Efficiency**: The significant reduction in false positives allows security analysts to focus on genuine threats, improving operational efficiency and reducing alert fatigue.
2. **Cost Optimization**: While the initial implementation of a big data infrastructure requires investment, the improved storage efficiency (3.8M events/GB vs. 1.2M events/GB for traditional SIEM) and better hardware utilization can lead to lower total cost of ownership over time.
3. **Adaptive Security**: The AI components enable continuous learning from new attack patterns, making the system more resilient against evolving threats without requiring constant manual rule updates.

4. **Threat Hunting**: The big data architecture supports interactive exploration of historical security data at scale, enabling proactive threat hunting that would be impractical with traditional SIEM systems.
5. **Compliance Requirements**: The improved detection capabilities and comprehensive data retention help organizations meet increasingly stringent regulatory requirements for security monitoring and incident response.

5.3 Limitations and Challenges

Despite the promising results, several limitations and challenges should be acknowledged:

1. **Data Quality Dependencies**: The effectiveness of machine learning models heavily depends on the quality and representativeness of the training data. Organizations with limited historical security data may face challenges during initial deployment.
2. **Interpretability Concerns**: Some of the advanced AI techniques, particularly deep learning models, operate as "black boxes," making it difficult to explain detection decisions to stakeholders or compliance auditors.
3. **Skills Gap**: Implementing and maintaining AI-enhanced SIEM systems requires specialized skills in both cybersecurity and data science, which may be challenging for organizations facing talent shortages.
4. **Adversarial Resilience**: Sophisticated attackers might attempt to evade machine learning detection through adversarial techniques, requiring ongoing research into more robust algorithms.
5. **Privacy Considerations**: The comprehensive data collection necessary for effective analysis raises privacy concerns, particularly in regions with strict data protection regulations like GDPR.

6. Conclusion and Future Work

6.1 Summary of Contributions

This research presented a comprehensive architecture for big data-driven SIEM enhanced by artificial intelligence. Our system integrates distributed computing technologies with advanced machine learning techniques to overcome the limitations of traditional SIEM solutions. The experimental evaluation demonstrated significant improvements in detection accuracy, processing performance, and scalability. Key contributions include:

1. A scalable, three-layer architecture for security event processing that effectively handles the volume, velocity, and variety challenges of big data
2. Novel applications of machine learning techniques for security event analysis, with particular emphasis on reducing false positives
3. Empirical evidence of the effectiveness of AI enhancement across diverse attack scenarios
4. Practical insights for implementing big data-driven SIEM systems in enterprise environments

6.2 Future Research Directions

Several promising directions for future research emerge from this work:

1. **Explainable AI for Security**: Developing techniques to improve the interpretability of complex machine learning models used in security detection, allowing analysts to understand why specific alerts were generated.
2. **Adversarial Machine Learning**: Investigating techniques to make security analytics more resilient against adversarial attacks that attempt to evade detection by manipulating input data.
3. **Transfer Learning for Security**: Exploring transfer learning approaches to address the cold-start problem for organizations with limited labeled security data.
4. **Automated Response Integration**: Extending the system to incorporate automated response capabilities that can take remedial actions based on high-confidence detections.
5. **Federated Learning for Security**: Investigating privacy-preserving federated learning techniques that allow multiple organizations to collaboratively improve detection models without sharing sensitive security data.
6. **Human-AI Collaboration**: Developing interactive interfaces that facilitate effective collaboration between security analysts and AI systems, combining human expertise with computational capabilities.

In conclusion, our research demonstrates that the integration of big data technologies and artificial intelligence can significantly enhance SIEM capabilities, enabling organizations to detect and respond to security threats more effectively in increasingly complex environments. While challenges remain, the approach presented here offers a promising path forward for next-generation security monitoring systems.

References

1. Abdel-Basset, M., Gamal, A., Chakrabortty, R. K., & Ryan, M. (2021). A new hybrid multi-criteria decision-making approach for location selection of distribution centers in supply chain network. Computers & Industrial Engineering, 157, 107324.
2. Alom, M. Z., & Taha, T. M. (2017). Network intrusion detection for cyber security using unsupervised deep learning approaches. In 2017 IEEE National Aerospace and Electronics Conference (NAECON) (pp. 63-69). IEEE.

3. Apruzzese, G., Colajanni, M., Ferretti, L., Guido, A., & Marchetti, M. (2018). On the effectiveness of machine and deep learning for cyber security. In 2018 10th International Conference on Cyber Conflict (CyCon) (pp. 371-390). IEEE.
4. Barros, A., & Chuvakin, A. (2015). How to build and operate a SIEM in today's age. Gartner Technical Professional Advice Report.
5. Bhatt, S., Manadhata, P. K., & Zomlot, L. (2018). The operational role of security information and event management systems. IEEE Security & Privacy, 12(5), 35-41.
6. Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications Surveys & Tutorials, 18(2), 1153-1176.
7. Huang, K., Zhou, C., Tian, Y. C., Tu, W., & Peng, Y. (2020). Application of bayesian network to data-driven cybersecurity risk assessment in SCADA networks. IEEE Transactions on Automation Science and Engineering, 17(3), 1376-1391.
8. Jeong, J. P., Kwon, E., Hwang, M. J., Seo, S. H., & Kim, B. J. (2019). A reliable Hadoop-based distributed monitoring system for detecting security attacks in real time. Computers & Security, 83, 353-366.
9. Kent, A. D. (2016). Comprehensive, multi-source cyber-security events. Los Alamos National Laboratory.
10. Leszczyna, R. (2019). A review of standards with cybersecurity requirements for smart grid. Computers & Security, 77, 262-276.
11. Miloslavskaya, N., & Tolstoy, A. (2016). Big Data, Fast Data and Data Lake Concepts. Procedia Computer Science, 88, 300-305.
12. Moustafa, N., & Slay, J. (2015). UNSW-NB15: A comprehensive data set for network intrusion detection systems. In 2015 Military Communications and Information Systems Conference (MilCIS) (pp. 1-6). IEEE.
13. Nicolett, M., & Kavanagh, K. M. (2009). Magic quadrant for security information and event management. Gartner RAS Core Research Note.
14. Sarker, I. H., Kayes, A. S. M., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity data science: an overview from machine learning perspective. Journal of Big Data, 7(1), 1-29.
15. Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. In Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP) (pp. 108-116).
16. Terzi, D. S., Terzi, R., & Sagiroglu, S. (2017). Big data analytics for network anomaly detection from netflow data. In 2017 International Conference on Computer Science and Engineering (UBMK) (pp. 592-597). IEEE.
17. Vinayakumar, R., Alazab, M., Soman, K. P., Poornachandran, P., Al-Nemrat, A., & Venkatraman, S. (2019). Deep learning approach for intelligent intrusion detection system. IEEE Access, 7, 41525-41550.
18. Zhao, Y., Li, Y., Zhang, X., Geng, G., Zhang, W., & Sun, Y. (2020). A survey of networking applications applying the software defined networking concept based on machine learning. IEEE Access, 8, 29344-29384.
19. Zuech, R., Khoshgoftaar, T. M., & Wald, R. (2015). Intrusion detection and big heterogeneous data: a survey. Journal of Big Data, 2(1), 1-41.