👽 Allen

# The Operational Role of Security Information and Event Management Systems

# [Allen Jose](mailto:allenjose2110@gmail.com) Sep 21, 2025 8:15 AM GMT+5:30 [snp14.pdf](https://drive.google.com/open?id=1gE2jmjb90P-PypHDAffOwnbuifzmL3DB) Completed

SIEM - Security Information and Event Management (Tool)

SOCs - Security Operations Center (Place or a Center)

CSIRTs - Computer Security Incident Response Teams (Team)

[**CSIRTs**](https://csirtsnetwork.eu/)receives, reviews, and responds to security threats and incidents, **SOCs** are the ones who monitor the firewalls, intrusion prevention devices, application servers, databases, and even the user accounts for security risks. **SIEM** are the tools which they use to collect, normalize, store, and correlate events to identify the malicious activities that’s happening in the enterprise IT assets.

**The Main use of SIEM tools:**

In the early days there were only a few systems available to assess the threats which includes the perimeter defence systems like firewall, intrusion detection system and Anti-Virus softwares for the host systems. As the time passed many such systems evolved and came into the market. Two main problems arose due to this situation and they were:

1. There were too many user Interfaces to manage
2. **There was no one such system that could correlate events across the security tools.**

Since there was no correlation these security tools had no idea about the system architecture which then triggered false positives for the system vulnerability.

This is exactly why the SIEM tools are made to meet these challenges. They correlate with different security tools, collect data from them, normalise those collected data using the vendor specific schema, and trigger the alerts only when a certain event gets into action which is stored in their **rule engine.**

Main Strength of the SIEM tools is to correlate the logs from different tools in order to define a meaningful attack pattern and scenarios which when happens alerts the Security Analyst (SAs). So in short they act like a radar system for the security breach. Their long term event retention capability can also help in the forensic analysis and detect slow attacks including the Advanced Persistence Threats (APTs).

Since the SIEM tools collect the logs and data from different sources their primary or the **First Task is to normalise (***from different representation to a common format***) the data that they have received for further processing, analysis and rule creation.**

***Specialised Connectors*** are the ones which parse the different input logs into a common format.

Once the data or the logs are normalised they are pushed into 2 different systems:

1. Security Management System
2. Forensic Analysis database

This platform stores and analyzes a window of events typically lasting a few hours. And if needed the database can even store them for a longer period of time for like 3 to 6 months. The injection rates also differ for the database and the management system, they are like 100 Kbyte per second for the database, and 10 Kbyte for the security management system. Enhanced hardware can boost these numbers further more.

The Security Management applies its rules into the window periodically and when any threats are being popped up they are being sent to the SIEM terminal for the SOCs analysis.

Each rule captures a different malicious behaviours for example:

1. Multiple failed login attempts in a window
2. Continuous request to access a block HTTP request.
3. Finished package sent to check whether a port is free or not.

Rules can be created or generated from 2 different sources, either the analyst her/herself or by the SIEM algorithm itself from the events by identifying the set of patterns that happens in the window of events.

The security management system is being overlooked by the SAs and there are multiple levels of SAs in a SOC. The level I SAs typically has to manage around 1000 to 3000 malicious alerts or threats detected by the SIEMs per day, and this number is completely dependent on the network traffic and load that it receives on that day. The level I SAs due to this heavy load of work they have to manage in 8 to 12 hours shifts. Because of this they tend to spend only 1 to 2 years in their jobs including their training periods. Their job is to classify whether the threat raised by the SIEM is an actual thread or a false alarm. If they find it in neither of them they tend to escalate this matter to the level 2 SAs. And if the SAs -II find the alert raised by the SA-I as a false positive they then work with the security analysts to find out the rule that triggered that false alarm in the first case or else if they find it as the actual threat they tend to escalate it to the SA-III.

The next few passages define the importance of the SIEM tools in the SOCs of the organizations which have a higher number of applications, devices, users. The total estimated cost to run a SIEM with initial budget can lead anywhere from $3 to $5 million and it also requires a higher number of manpower to work and manage it. So even a minimum of 20 SAs can shoot up the cost into $5 million again.

**The main challenges that is faced by the SAs in the SOCs:**

**Lack of Contextual Information**: Since the SAs are a part of design, development, or testing they find it difficult to relate the logs that come from different devices. Routine actions like patching, backup, testing might trigger alerts in the SIEMs.

Since the SOCs monitor only a particular length of window in the event log systems they are unable to track stealthy and slow advancing attacks in the systems especially the APTs. To address this issue they try to sample a few raw logs in order to find an attack pattern to detect the stealthy attacks like APTs.

Scalability is another issue with the SIEM in the SOCs. For instance, if we need to log **1 trillion events into our SIEM we need to have an injection rate of 12 million events** into the SIEM per second. These values are several fold higher to the state of the art machineries that we have today. The values tend to rise proportionally to the complexity of the devices and the number of employees that an organization has.

The **next Big Challenge that a SIEM faces is** to decide which event or the **right even**t to collect from the devices as there are a million more events that happen. Collecting and correlating the right events is the most important method to solve a security problem.

The next hurdle a SIEM has is the storage limitations. For example as mentioned in the paper if 1 event requires 400 bytes of space and if we achieve 10x data compression, then for a 1 trillion events we might need 40 Tbytes of storage in the forensic analysis database.

Sep 21, 2025 12:45 AM

# A Security Information and Event Management Pattern

[Allen Jose](mailto:allenjose2110@gmail.com) Sep 21, 2025 2:00 PM [SIEM\_Pattern.pdf](https://drive.google.com/open?id=1iALDjM1tgH460BED69otHIOWwFOAJ5Ih) Completed

This paper focuses on laying out a basic understanding on the SIEM components and they proposed a pattern for the SIEM. They are also claiming that the proposed pattern will certainly help the software developers to clearly define the boundaries and functions of different SIEM components which enables the enhancement of reusability and replaceability of the selected modules. The pattern was deduced from the [*DINGfest*](https://www.researchgate.net/figure/DINGfest-conceptual-architecture-and-data-flow_fig1_325193079) project which is funded and developed by the Federal Ministry of Education and Research in Germany, and their primary aim is to collect the data for forensic analysis from different virtual environments, furthermore they also focused on Visual Security Analysis.

The First notion of the SIEM systems is attributed to a report by Garter Inc by Williams and Nicolet in 2005. At first the SIEM was two independent systems like SIM and SEM. Where the SIM is called Security Information Management, and the SEM is called Security Event Management. While SIM deals with centralized management, collection, preservation of historic log data and the generation of reports for compliance purposes, SEM covers threat management, real-time monitoring of security incidents and triggering proper reactions in case of an incident. SIM and SEM serve distinct but complementary roles in security management. SIM focuses on centralized management, collecting and preserving historical log data, and generating compliance reports. SEM, on the other hand, is dedicated to threat management, real-time monitoring of security incidents, and initiating appropriate responses when an incident occurs.

There are several problems that a SIEM system faces even with its complicated and sophisticated systems. Few of them which were mentioned in the paper includes:

1. Detection rate and False Positives
2. Time of discovery and reaction
3. Useability
4. Visual preparation of data and integration of expert knowledge
5. Degree of automation
6. Number of connectible devices
7. Analytics
8. Reusability and intelligence sharing
9. Costs.

In the paper they have also proposed an architecture which solves the problems the main SIEM pattern was split into 8 components

First Module: Responsible for log collection from different modules and providing the collected data to the further modules. The input can be from various devices.

Second Module: Enrichment component uses the context data for providing additional information for the bare log data in order to improve the further processing by providing the enriched log data. Various context data sources supply the needed data.

Third Module: For the further processing the logs collected from different devices are being restructured to standardize format, in simple words the data are going under the normalization component.

Fourth Module: Correlation and analysis component. Here the incidents are finally recognized based on the incoming normalized data. In addition the analysis are conducted based on the historic data provided by the storage component.

Fifth Module: It is the storage component where the logs are being stored for future analysis in order to track the stealth attack like APTs.

Sixth Module: Reporting Module, if the fourth module i.e., the correlation and the analysis component found any event that is even as a slight threat or attack it’ll forward the issue to this module in order to give the alert to the SAs in the SOCs.

Seventh Module: Monitoring Component, it provides an interface for the human interaction and the alerting & incident response component.

Eighth Module: Alerting & Response Module, this component can propagate the alerts to the responsible stakeholders or react automatically in case of an incident.

Compliance and legal frameworks drive reporting — in regions like the EU and USA, incidents often must be formally reported, and SIEM-generated reports help meet those requirements.

Comparison with IBM QRadar shows how the abstract SIEM pattern maps onto real industry tools, though vendors brand their components differently (e.g., QRadar’s collectors, risk managers, custom rules engine).

The DINGfest research project extends the basic SIEM pattern with unique features such as pseudonymization for privacy, snapshotting whole VMs for forensics, and advanced visual analytics for human experts. Unlike traditional SIEMs, it prioritizes privacy by pseudonymizing all stored and analyzed data, with depseudonymization only allowed on strong suspicion.

Some SIEM systems invert the order of enrichment and normalization — normalizing first, then enriching. Others also outsource parts of the pattern (e.g., storage) to cloud services without changing the overall structure.

Variants like “active defense” SIEMs extend beyond monitoring to mitigate or stop attacks proactively, relying on centralized threat intelligence feeds.

Examples of well-known SIEMs fitting this pattern include Splunk (extended with UEBA), LogRhythm (with AI-assisted monitoring), McAfee ESM (integrating big data tech like Elasticsearch), and AlienVault OSSIM/USM (popular for its open-source base).

Advantages of the SIEM pattern include better detection rates, reduced false positives through correlated/contextual analysis, faster response times due to centralized monitoring, and higher automation for incident response and compliance reporting.

Limitations include the high effort of integrating SIEM into complex, historically grown infrastructures, the continuing need for expert analysts despite automation, and the lack of mechanisms for incorporating human-sourced intelligence (like a phishing phone call reported by an employee).

# 

# Security Information and Event Management Analysis, Trends, and Usage in Critical Infrastructure

[Allen Jose](mailto:allenjose2110@gmail.com) Sep 22, 2025 9:30 AM [SIEM\_A,T,UI.pdf](https://drive.google.com/open?id=15r13rK3czYemvQwwX1dFC-161lMn85Eq) Completed

The role of SIEM systems in industrial and IT infrastructures is emphasized not only as a monitoring and correlation platform but as a technology that has steadily evolved into a comprehensive defense mechanism. What distinguishes this evolution from earlier stages is the integration of real-time anomaly detection, faster incident management, and advanced visualization tools. These features, once considered add-ons, are now seen as baseline requirements for SIEMs, especially in the context of industrial control systems where attackers can cause not just data loss but physical and operational disruption. The sophistication of attackers has reached a point where they are often backed by organized groups or state-level actors, which raises the stakes for any defensive system. This also creates the need for SIEMs to combine traditional log-based detection with behavioral analysis to capture subtler anomalies that may precede a major attack.

Another important distinction that this work highlights is the gradual overlap between SIEMs and big data analytics. The explosion of networked devices and applications, combined with the surge of data from IT and OT environments, requires platforms that are no longer confined to classical log correlation. SIEMs are moving into the territory of large-scale analytics platforms, leveraging high-volume event processing to identify risk areas across entire organizations. This convergence reflects a trend where SIEMs are not only tools for threat detection but also enablers of proactive risk management strategies, helping organizations reduce costs and accelerate their response times.

The landscape of SIEM vendors has been shaped over a decade of classification by organizations like Gartner, who divide them into leaders, challengers, niche players, and visionaries. Leaders such as IBM’s QRadar, RSA’s NetWitness, and MicroFocus ArcSight have maintained long-term positions in the market by executing consistently well. Visionaries such as AlienVault have contributed innovative features, particularly around open-source accessibility and integrated monitoring. Meanwhile, platforms like Splunk have introduced operational intelligence as a core feature, making SIEM less of a static monitoring tool and more of an exploratory environment where analysts can mine vast amounts of data for insights. This categorization underscores that not all SIEMs are built with the same philosophy—some prioritize compliance and log management, while others aim to be analytics-first platforms.

When considering the features of modern SIEMs, there are both strengths and persistent gaps. All major SIEMs provide real-time event processing and correlation, but only a few go beyond simple rule-based engines. Some, like Splunk, incorporate machine learning and behavioral analytics, while others remain dependent on manual rule creation. Visualization is another area where significant differences appear—many systems still offer rigid dashboards that are poorly suited for interactive exploration, limiting the analyst’s ability to connect patterns across different data sources. This lack of advanced visualization often forces analysts to rely on their own scripts or third-party tools, which increases the operational complexity of SOC environments.

The scalability of SIEMs is another recurring concern. With the growth of connected devices and digital transformation initiatives, the volume of logs and events has increased exponentially. While modern systems claim high scalability, most of them still struggle with maintaining performance when handling truly massive datasets. For example, while correlation engines are capable of near real-time detection, the indexing and search capabilities often degrade when faced with months of archived data. As a result, forensic investigations can be slow and incomplete, especially when advanced persistent threats unfold over longer periods. This limitation makes it clear that next-generation SIEMs need to rethink storage and retrieval, potentially by integrating distributed storage systems or cloud-based archives that allow elastic retention without sacrificing access speed.

From an operational perspective, one of the major limitations noted in current SIEMs is their dependence on human analysts for decision-making. Even though automation is being gradually introduced through security orchestration and response mechanisms, most SIEMs still rely heavily on analysts to interpret alerts, decide on countermeasures, and implement them. This human dependency introduces delays and potential errors, particularly in high-volume environments where analysts face thousands of alerts per day. Automating response mechanisms is seen as a necessary step, but it also raises challenges around accuracy and impact analysis—deploying a countermeasure without understanding its potential side effects can cause more disruption than the attack itself.

Another unique challenge addressed in the study is the issue of incomplete data. While SIEMs collect vast amounts of logs, they rarely capture all the necessary data to build a complete picture of user or system activity. Critical data types like DNS traffic, endpoint behavior, and email logs are often missing from correlation engines. Without these, analysts are left with fragmented views, which results in both false positives and undetected attacks. Moreover, identity remains a fragmented concept across SIEM systems—shared accounts and role overlaps mean that linking an activity to a specific individual can be legally and technically problematic, especially under strict regulations like the GDPR. This introduces an additional dimension: SIEMs must balance privacy requirements with the need to provide enough visibility for accurate threat detection.

Looking at the future of SIEMs, external factors play a decisive role. Political initiatives, such as the European Union’s investments in cybersecurity, can accelerate the development of new SIEM capabilities by creating both funding and regulatory pressure. Economically, the rise of freelance and short-term employment models means that organizations increasingly rely on external devices and contractors, which introduces monitoring challenges for SIEM platforms. From a societal perspective, the spread of Generation Z into the workforce implies greater awareness of cybersecurity but also heavier use of social networks, which themselves become valuable sources of security-relevant data. Social media analytics, once outside the scope of SIEM, are gradually being considered as inputs for threat intelligence, particularly for early detection of phishing campaigns, exploit discussions, or coordinated attacks.

Technological factors are perhaps the most influential. The rise of cloud services has transformed both storage and application deployment models, forcing SIEMs to integrate with hybrid and multi-cloud environments. Mobile and BYOD practices further expand the attack surface, introducing devices that are often outside the control of corporate IT. Big data analytics and machine learning are seen as the natural enablers of the next generation of SIEMs, allowing detection engines to move beyond static rules into dynamic behavior modeling. The emergence of the Internet of Everything and 5G networks compounds the challenge by dramatically increasing both the volume and velocity of data. SIEM platforms will need to handle not only traditional enterprise logs but also sensor data, IoT activity, and edge communications at unprecedented scale.

Legal and regulatory factors also reshape the SIEM landscape. With data protection regulations like GDPR classifying IP addresses and other identifiers as personal data, SIEMs must redesign how they collect, store, and share information. This introduces constraints on where data can be stored, how long it can be retained, and what anonymization measures must be applied. Such legal pressures mean that SIEM vendors cannot rely solely on technical efficiency—they must also ensure compliance, which complicates storage architectures and cross-border data flows. Environmental factors, particularly the spread of IoT and cloud adoption, add to this challenge by producing massive streams of data from devices that are often insecure by design, creating fertile ground for attackers.

To address these challenges, the paper highlights several enhancements for the next generation of SIEMs. One is the integration of open-source intelligence (OSINT) directly into SIEM workflows, allowing real-time tagging of threat-related information from external sources such as forums or social networks. Another is the development of more advanced visualization models capable of handling streaming data, high-dimensional datasets, and temporal patterns. Enhanced storage mechanisms, particularly elastic and secure cloud-based archives, are proposed as solutions to the long-standing retention problem. Perhaps most significantly, the convergence of SIEM with SOAR technologies is seen as inevitable, creating platforms that combine orchestration, automation, and incident response into a unified system.

The paper also emphasizes the potential of AI and machine learning integration. While a few SIEMs already embed machine learning engines, their use remains limited. The vision for the next generation is a SIEM that continuously learns normal patterns of user and application behavior, flags deviations, and even adjusts its own detection rules accordingly. This would reduce the reliance on human analysts, cut down false positives, and enable faster response times. A hybrid model combining rule-based detection with AI-driven behavioral analysis is considered the most promising approach, as it combines the transparency of rules with the adaptability of machine learning.

Finally, the application of SIEMs in critical infrastructures reveals unique challenges. Industrial systems prioritize availability over confidentiality, meaning they cannot tolerate downtime caused by aggressive scanning or intrusive detection methods. These systems often run on outdated or fragile platforms, making them harder to secure. The integration of SIEMs into such environments requires adaptations, such as non-intrusive monitoring and tailored correlation rules that account for operational constraints. Examples in energy, water, transportation, healthcare, and finance illustrate how SIEMs can provide visibility into both cyber and physical dimensions of attacks. For instance, in power grids, SIEMs can detect anomalies like GPS spoofing or denial-of-service attempts on smart meters. In water supply, they can correlate changes in sensor readings with suspicious network behavior to detect tampering. In healthcare, SIEMs help organizations stay compliant with HIPAA while monitoring the growing number of personal medical devices connected to networks. In finance, they play a role not only in detecting fraud but also in generating mandatory incident reports required by regulatory authorities.

# Challenges and Directions in Security Information and Event Management System.

[Allen Jose](mailto:allenjose2110@gmail.com) Sep 23, 2025 3:00 PM [C&D\_SIEM.pdf](https://drive.google.com/open?id=1CxZIF36T4VGGB7wDVVzsheI8LFprhh3x) Completed

The integration of SIEM into complex industrial systems is not as seamless as often assumed. While SIEMs are widely deployed in Security Operation Centers to collect and normalize data for analysis, the actual deployment process tends to rely heavily on ad-hoc connectors and handcrafted correlation rules. This dependence means that analysts must anticipate threats in advance, which is a flawed assumption given the dynamic and unpredictable nature of attacks. In practice, the system described in this work—an Air Traffic Control (ATC) platform—produces massive volumes of unstructured proprietary logs, far removed from the structured input formats that commercial SIEMs are optimized to handle. This makes the reliance on correlation rules even less effective, since meaningful threat patterns cannot always be predefined by humans.

The ATC system studied here consists of multiple distributed nodes carrying out mission-critical functions such as flight progress monitoring and conflict detection. Together they generate more than 15,000 log lines per minute, with bursts where certain logs emit over 3,000 lines within just 10 seconds. Importantly, these logs are unstructured and are often written for human operators rather than machines, including verbose warning messages and descriptive outputs. Unlike standard syslogs, which have well-defined fields such as timestamp and severity, these application logs resist automated parsing. Commercial SIEM frameworks like Splunk or LogRhythm provide internal models to standardize structured inputs, but they perform poorly when fed with unstructured free text. Consequently, a large portion of the log data is ingested in raw form, making automated correlation less reliable.

Another difficulty raised by this work is the absence of a reliable ground truth in security data. Training supervised models for anomaly detection requires labeled incidents, but in production systems such labels are rare, costly to obtain, and incomplete. Even if labeling were possible, attacks evolve constantly, producing entirely new patterns that were not present in historical data. This dynamic nature leads to highly imbalanced datasets where anomalies are only a fraction of the recorded events. Thus, conventional supervised learning becomes impractical in domains such as ATC, where only a handful of true incidents may occur over years of operation.

The authors explore the use of OSSEC, an open-source monitoring solution that combines HIDS and log analysis, to integrate with the ATC system. Since OSSEC relies on syslog input, a connector was implemented to wrap the unstructured ATC logs into syslog format. This workaround allowed the system to reuse OSSEC’s existing parsing and alerting mechanisms while extending them with custom rules. A simple proof-of-concept showed how domain-specific keywords could be used to trigger alerts. However, this still depended on manually writing XML rule files, which illustrates the fundamental problem: analysts are expected to know in advance what suspicious behavior will look like, which is rarely the case in evolving environments.

To address these limitations, the study introduces behavioral baselines as a more robust approach. Instead of trying to predict every possible attack signature, the idea is to model what constitutes normal behavior and then flag deviations from this baseline. Because the ATC logs are largely textual and unstructured, the authors turn to natural language processing methods, specifically Latent Dirichlet Allocation (LDA), to extract patterns from the logs. Here, chunks of log data collected over 10-second intervals are treated as documents, and LDA is used to identify recurring topics that represent regular system actions. The result is a probabilistic model of normal operations without requiring labeled incidents or pre-defined correlation rules.

The evaluation showed that LDA models can distinguish between regular and anomalous log sequences with promising accuracy. By artificially injecting anomalous lines into log chunks, the authors measured how perplexity scores diverged from the baseline model. Even with as little as 1% of anomalous content, the model was able to detect deviations with high recall and precision. At higher levels of anomaly, the separation from regular data was clear, with virtually no false positives or negatives. Unlike rule-based systems, this approach requires no prior threat knowledge, no handcrafted indicators, and no supervision. Instead, it leverages the structure inherent in the logs themselves to learn what is normal and to highlight departures.

The significance of this approach is twofold. First, it provides a path toward automated data discovery in SIEM, reducing the dependency on human analysts who are otherwise tasked with writing thousands of rules. Second, it demonstrates that unsupervised machine learning methods can be adapted to highly specialized industrial domains, such as ATC, where threat models are immature and attack patterns are not yet well understood. In these contexts, baseline modeling is more practical than attempting exhaustive rule coverage.

The work also highlights the broader research directions needed for SIEM evolution. Future SIEM platforms should integrate plugins or modular extensions that can handle domain-specific log formats without requiring heavy customization. Visualization remains another open area, since analysts need ways to interpret and explore the learned baselines interactively. Finally, the long-term goal is to design proof-of-concept automatons that can mimic at least part of the reasoning carried out by human analysts, thereby reducing the cognitive workload in high-volume environments like SOCs.

In conclusion, this study positions SIEM not only as a central element of SOCs but as a practice in need of augmentation. The reliance on rules and structured data is a limiting factor in domains characterized by unstructured, high-volume logs and scarce ground truth. By combining traditional monitoring with data discovery techniques like LDA, SIEM can evolve into a more adaptive platform capable of detecting anomalies without prior assumptions. This direction represents a shift from rule-driven detection toward learning-driven discovery, which may prove critical in mission-critical infrastructures where unknown threats are the most dangerous.

🧑‍💻 Satish

**▶️Refer Sub Tabs**

1️⃣ Analysis, Trends, and Usage in Critical

Security Information and Event Management (SIEM): Analysis, Trends, and Usage in Critical Infrastructures

[Satish Pakalapati](mailto:satishpakalapati65@gmail.com) 20 Sept 2025 08:00 [sensors-21-04759-with-cover.pdf](https://drive.google.com/open?id=15r13rK3czYemvQwwX1dFC-161lMn85Eq) Completed

Keywords for further research :

Evolution of SIEMs;

SIEM enhancement;

SIEM trends;

Critical infrastructures

Abstract Highlights:

* Security Information and Event Management (SIEM) systems have been widely deployed as a powerful tool to prevent, detect, and react against cyber-attacks.
* comprehensive systems that provide a wide visibility to identify areas of high risks and proactively focus on mitigation strategies aiming at reducing costs and time for incident response.
* converging with big data analytics tools.

Intro:

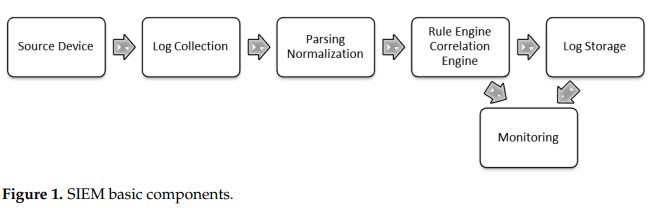
current cybersecurity incidents affecting IT and ICS are

* Cyber Threats: Organizations face risks such as ransomware, malware disrupting operations, phishing targeting executives/IT staff, business email compromise, data theft, and social engineering.
* NIST Guidelines: Cybersecurity in industrial control systems should enable real-time anomaly detection, faster incident management, and intelligent network visualization.
* Role of SIEMs: SIEMs act as the core of security operations centers by collecting and correlating events from sources like IDS, firewalls, and antivirus. Key vendors include HP, IBM, Intel, McAfee, AT&T Cybersecurity/AlienVault, and Splunk.

SIEM Solutions

1. Architecture:

SIEM is composed of separate blocks (e.g., source device, log collection, parsing normalization, rule engine, log storage, event monitoring)



1. SIEM Classification :

While some research focuses on the commercial aspects, others concentrate on the technical features that could be improved in current SIEM solutions.

Based on Gartner’s criteria.

1. Leaders: Execute current vision well and are well positioned for the future.
2. Visionaries: Have strong future vision but weak execution today.
3. Niche Players: Focus on small segments or lack innovation compared to others.
4. Challengers: Execute strongly now or dominate a segment, but lack clear future direction.

The evolution of SIEM solutions (and SIEM vendors) from 2010 to 2020

★ = Leaders

◆ = Challengers

▲ = Niche Players

■ = Visionaries

| Category | Vendors / Notes |
| --- | --- |
| Consistently in Top Rankings (2010–2020) | RSA (NetWitness ★), IBM (QRadar ★), ArcSight – NetIQ/MicroFocus/HP (★◆), McAfee (Enterprise Security Manager ★), LogRhythm (NextGen SIEM ★) |
| Mergers & Evolution | IBM + Q1 Labs (2012–2013 ★), NetIQ + Novell (2012 ◆), HP + ArcSight (2013 ★), AccelOps + Fortinet (2016 ▲), Micro Focus + ArcSight & NetIQ (2017 ◆) |
| Always in Gartner Classification Since Entry | Splunk (★), AlienVault/AT&T Cybersecurity (■), SolarWinds (▲), EventTracker (▲), Fortinet (◆/▲), MicroFocus (◆) |
| Recent Entrants (with UEBA features) | ManageEngine (▲), Venustech (▲), Rapid7 (■), Exabeam (■), Secureonix (■), LogPoint  (▲), HanSight (■) |

Table of classification based on Gartner’s criteria.

Consistent Leaders: Only a few vendors (IBM QRadar, RSA NetWitness, ArcSight, McAfee, LogRhythm) stayed in Gartner’s top rankings across the decade.

Market Shaped by Mergers: Major consolidations (IBM–Q1 Labs, HP–ArcSight, MicroFocus–NetIQ) significantly influenced SIEM evolution.

Shift to Advanced Analytics: Recent entrants (Exabeam, Secureonix, Rapid7, etc.) with UEBA features highlight the trend toward behavior-based detection.

Each Tool has its own limitations

3. SIEM Features and Capabilities :

Features:

* Correlation Rules: A SIEM's core function. Advanced systems use powerful rules to detect complex threats.
* Data Sources: The more native data sources a SIEM supports, the better its visibility.
* Real-time Processing: Must handle millions of events instantly for immediate threat detection.
* Data Volume: Its capacity to store and analyze large amounts of data.
* Visualization: Clear dashboards are vital for security analysts to quickly understand events.
* Data Analytics: Advanced SIEMs use machine learning (UEBA) to spot abnormal behavior.
* Performance: A measure of its speed and efficiency in processing, searching, and storing data.
* Forensics: Its ability to capture and analyze network data for investigations.

Capabilities:

* Complexity: How difficult it is to deploy and manage.
* Scalability: Its ability to grow with an organization without losing performance.
* Risk Analysis: Its capacity to assess risks on assets.
* Storage: How long it can retain data for analysis.
* Resilience: Its ability to handle failures and maintain continuous operation.
* Reaction & Reporting: Its built-in tools for responding to incidents and creating reports.
* Security: Features like automation and encryption to protect data within the system.
* UEBA: A modern feature that detects anomalies in user behavior.

### 4. Limitations of Current SIEMs

### Incomplete Data and Poor Context

Current SIEMs don't have all the data needed for comprehensive threat detection. They often miss crucial information from sources like DNS traffic, endpoint logs, and email. This results in an incomplete picture of network activity, leading to high rates of false positives and negatives.

* Privacy Issues: SIEMs also struggle with fragmented user identities and shared accounts, and privacy regulations like GDPR complicate the process of correlating this data. Future SIEMs need to balance providing actionable intelligence with respecting user privacy.
* Lack of Metrics: Existing SIEMs fail to provide high-level, quantitative security risk metrics. Security Operations Centers (SOCs) need measurable evidence to make informed decisions, but current systems don't provide this, especially with regard to cost-sensitive metrics.

### 2. Basic Correlation Rules

Most SIEMs rely on basic boolean logic and simple event chaining to detect threats. This is insufficient for catching multi-stage, sophisticated attacks like zero-day exploits. The focus on syntax over semantics in correlation languages means they lack the advanced capabilities needed to perform historical or behavioral analysis.

### 3. Basic Storage Capabilities

Current SIEMs have limited "live" storage, typically retaining data for only about 90 days. Once data is archived, it's difficult to access and analyze, hindering investigations into advanced persistent threats (APTs) that can go undetected for months. The process of archiving data is often manual and custom-scripted, which is inefficient, costly, and can introduce security and reliability risks.

* Future Goal: Future SIEMs should offer secure and elastic solutions for long-term data archival, regardless of retention policies, and make archived data easily searchable.

4. Reliance on Humans

Despite advancements in automation, threat analysis and response are still heavily dependent on human expertise. This process is slow, costly, and prone to error. While some SIEMs can trigger basic automated actions like sending an email or executing a script, they lack pre-configured, customizable, and comprehensive automated response capabilities.

### 5. Limited Reporting and Visualization

Current SIEMs often provide generic reports and visualizations that are not tailored to specific user needs, making it difficult to extract actionable insights from large datasets. They struggle to integrate and visualize data from diverse sources like statistical models or open-source intelligence. Future SIEMs need to provide flexible platforms with novel representations that help analysts understand attack origins, ongoing activities, and user behavior.

5. The Future of SIEMs:

The evolution of SIEM systems is not just driven by technology but also by external factors, which can be analyzed using a PESTLE (Political, Economic, Societal, Technological, Legal, and Environmental) framework.

1. Political Factors

Political initiatives and regulations play a significant role in shaping the SIEM market. For example, the EU is investing heavily in cybersecurity, with a planned investment of EUR 1.8 billion by 2020. This push for a secure "Digital Single Market" encourages the development and adoption of high-quality, interoperable cybersecurity products, including SIEMs, through public-private partnerships. This political backing provides a strong incentive for SIEM providers to innovate and expand their solutions to meet growing security demands.

2. Economic Factors

Several economic trends are influencing the future of SIEMs:

Changing Employment: The rise of short-term and freelance work means companies need tools that are easy for new employees to learn and use quickly. This makes user-friendly SIEM interfaces and simplified configuration crucial.

Freelance Consultants: Freelance cybersecurity consultants, who gain broad experience across different companies, can be a valuable resource for SIEM providers and users.

Job Growth: The cybersecurity job market is growing, with an estimated 35% growth by 2020, which creates a larger customer base for SIEM solutions.

Globalization: The growth of large, global companies and the increasing threat to small and medium-sized enterprises (SMEs) are driving demand for scalable and more affordable SIEM solutions, such as "SIEM as a Service."

3. Societal Factors

Societal changes are creating new challenges and opportunities for SIEMs:

"Generation Z": Younger generations, who are more digitally savvy, are expected to bring a higher awareness of cybersecurity risks to companies.

Social Media: The explosion of social media use creates a new, vast source of data that SIEMs must learn to analyze for security threats and event correlation.

Cyber-Attacks as "Weapons": The increasing use of cyber-attacks against critical infrastructure makes SIEMs essential for protecting not only companies but also citizens from physical and operational harm.

Deep Web: The existence of the deep web, which is not indexed by search engines, is a barrier that makes it difficult for SIEMs to retrieve data and intelligence.

4. Technological Factors

Technology is the primary enabler for SIEM evolution:

Cloud Computing: Cloud storage and services offer scalability and high availability for SIEMs, allowing them to handle massive amounts of data more efficiently and securely.

Mobile and BYOD: The proliferation of mobile devices and the "Bring Your Own Device" (BYOD) trend create new attack vectors and a need for SIEMs to monitor unmanaged devices and personal computers.

Big Data and Machine Learning: The exponential growth of data is pushing SIEMs to integrate advanced analytics and machine learning to make event detection and decision-making smarter and more efficient.

IoT and 5G: The Internet of Everything (IoE) and the increased data transfer speeds of 5G networks will generate a massive influx of data, requiring future SIEMs to adapt their collection and analysis capabilities.

Social Media Analytics: Hackers and security professionals are both using social media to share information about attacks, making it a critical source of intelligence that SIEMs must integrate.

5. Legal Factors

Legal regulations, particularly data protection laws, are profoundly impacting SIEM design:

GDPR: The EU's General Data Protection Regulation (GDPR) mandates strict rules on how personal data is processed and stored. This affects SIEMs because they often analyze personal data like IP addresses. As a result, future SIEMs must comply with these regulations by ensuring data is stored securely and processed in a way that respects privacy.

Location and Protection: Data protection laws influence where a SIEM's database can be located and require that the data is protected with an adequate level of security.

6. Environmental Factors

Environmental factors, in this context, refer to the broader technological environment:

Cloud Services: The increasing adoption of cloud services, especially by SMBs, drives the need for SIEMs that can deploy modules from the cloud and offer solutions like SIEM as a Service.

IoT: The sheer number of IoT devices creates a massive number of new endpoints and attack vectors. This forces SIEM systems to evolve to collect, organize, and analyze the influx of data from these diverse devices.

6. Potential Enhancements of Future SIEMs:

To move from manual, reactive security to an automated, proactive defense, SIEMs need to evolve in several key areas. The proposed enhancements focus on diversifying security data, improving analytics, and integrating new technologies.

1. Diverse Security

Future SIEMs must go beyond simple correlation and incorporate "diversity metrics." This means a SIEM should not only aggregate data from different sources but also understand how the strengths and weaknesses of various security controls (e.g., firewalls, IDSs) combine to protect the entire system. Instead of just aggregating diverse machine learning techniques, the focus should be on diverse inputs to build a more comprehensive and resilient defense.

2. OSINT Data Fusion

Future SIEMs should be able to process and integrate Open-Source Intelligence (OSINT) data. This could involve using natural language processing to identify threats from public sources like social media by searching for keywords like "DDoS" or "security breach." By fusing this external, real-time data with internal logs, SIEMs can provide a more complete picture of an evolving threat, helping to reduce false alarms.

3. Enhanced Visualization

Visualization is key to helping human analysts make sense of massive datasets. Future SIEMs need to:

Create specialized models for different data types (e.g., temporal, textual, spatial) to provide effective overviews and detailed views.

Support real-time analysis by handling the dynamic nature of streaming data from logs and OSINT sources.

Develop visual summaries of user activities to quickly identify abnormal or malicious patterns.

4. Enhanced Storage

Archiving and retrieving old data is a major challenge. The solution is to leverage public cloud services like Amazon S3 or Microsoft Azure. This would allow for a secure, elastic, and flexible data archival system, making it easier to store logs for long periods and conduct forensic analysis on advanced persistent threats (APTs) that might have gone unnoticed for months.

5. Integration with SOAR

The next generation of SIEMs must integrate with Security Orchestration, Automation, and Response (SOAR) platforms. While a SIEM is excellent at detecting threats and generating alerts, it still relies on human intervention for a response. SOAR automates these repetitive tasks, orchestrates workflows, and provides rapid incident response, allowing SIEMs to go beyond just alerting to actively containing and mitigating threats. This synergy allows security teams to focus on complex, strategic work rather than manual, routine tasks.

6. AI/ML Capabilities

AI and Machine Learning (AI/ML) should be at the core of future SIEMs. They can be used to:

Improve detection and correlation by identifying abnormal behavior in network traffic, users, and applications.

Reduce false positives by accurately classifying data as normal or abnormal.

Enable predictive capabilities by learning about threats as they acquire intelligence.

Automate responses and reduce dependency on manual intervention.

7. Other Enhancements

Risk-based Metrics: Develop high-level, quantitative security risk metrics that consider multiple layers of dependencies to provide evidence-based support for decision-making.

Hierarchical SIEMs: With the rise of 5G and IoT, future architectures may require a hierarchy of lighter, domain-specific SIEMs that can collaborate to process data closer to the network's edge.

Integration with XDR: SIEMs should be integrated with Extended Detection and Response (XDR) platforms. This allows the SIEM to focus on broad compliance and operational risk, while XDR provides a platform for deep, granular threat detection and response.

Increased Autonomy: To lower costs and simplify usage, future SIEMs must become more autonomous, requiring less effort for deployment and management.

7. SIEMs in Critical Infrastructures

Critical Infrastructures (CIs) are organizational and physical systems vital for public safety and security, such as energy, water, and healthcare. Unlike traditional IT networks which prioritize \*\*confidentiality and integrity\*\*, CI networks focus on \*\*availability\*\*, ensuring the system is always operational. This difference makes them highly vulnerable, as they often run on unpatched, fragile legacy systems that were not designed with security in mind. SIEM systems are an ideal solution to address these vulnerabilities, as they provide high-performance correlation and real-time event analysis from a business perspective.

SIEMs in Critical Infrastructure Sectors

SIEMs are being adapted and deployed in various critical sectors to provide crucial protection.

1. Energy Distribution

The energy sector is highly vulnerable due to the increasing number of threats and the complex, widespread nature of its infrastructure. SIEMs are being used to monitor signal strengths and identify anomalies in real-time, helping to detect and mitigate attacks like \*\*sleep deprivation, DDoS, and GPS spoofing\*\*. This technology is proving essential for protecting critical assets and ensuring the continuous supply of power.

2. Water Supply

The water sector is a prime target for cyber-attacks on treatment plants and distribution centers. SIEMs are used to monitor the entire \*\*SCADA network\*\* in real time, detecting deviations from acceptable water quality ranges. By correlating security events from the industrial control network, SIEMs can identify a wide range of attacks, from reconnaissance and malware to man-in-the-middle attacks. They are key to plugging security gaps and providing a holistic, real-time view of both IT and OT environments.

3. Transportation

As transportation networks become increasingly digitized, they face a growing number of vulnerabilities. SIEM systems are crucial for monitoring data flow and detecting threats that could lead to physical damage, service disruption, or data breaches. They help develop comprehensive cybersecurity plans, improve security in systems, and manage security in a proactive and reactive way to protect against risks like ransomware and cyber espionage.

4. Healthcare

The healthcare sector is a major target for cyber threats like ransomware and phishing. SIEMs are essential here for their ability to provide \*\*real-time analytics, scalable log management, and compliance reports\*\* (e.g., for \*\*HIPAA\*\*). SIEMs give security administrators a consolidated view of security events, which is invaluable for preventing data breaches and protecting sensitive patient information, especially with the growing use of medical IoT devices.

5. Financial Services

Financial organizations are a top target for insider and external threats. SIEMs are used to combat these threats through \*\*User and Entity Behavior Analysis (UEBA)\*\*, which identifies baseline user behaviors and detects any deviations that could signal an insider threat or data exfiltration. SIEMs also help financial institutions with \*\*regulatory compliance\*\* by automating the creation and filling of reports, a major challenge in this highly regulated sector. They are also vital for audit trail protection, forensics, and fraud detection.

### Current SIEM Limitations

* Commercial Focus: Market analyses from institutions like Gartner and TechTarget prioritize commercial aspects like market share and vendor comparisons over technical features and capabilities.
* Technical Gaps: Current SIEMs have weaknesses in core functions:
  + Visualization: They struggle to effectively visualize the massive volume of data they collect, hindering analysts' ability to make quick decisions.
  + Storage: On-premise storage is limited and expensive, while automated data archival is often manual and custom-scripted.
  + Automation: Their automated reaction and response capabilities are limited, often requiring extensive manual intervention, especially in critical infrastructure.

### The Future of SIEMs

* Evolution to Data Analytics: SIEMs are moving beyond basic log correlation to become advanced data analytics platforms.
* Holistic Approach: Future SIEMs need to be holistic, addressing technical and external factors simultaneously.
* Key Enhancements:
  + Diverse Sensor Frameworks: They must be able to integrate data from diverse and redundant sensors.
  + Risk-Based Metrics: They should provide multi-level, quantitative risk metrics to support evidence-based decisions.
  + Improved Visualization: They need better visualization and analysis tools to help analysts gain a high-level understanding of events.
  + Cloud Storage: Cloud-of-clouds storage is a promising solution for secure, elastic, and cost-effective long-term data archival.
* Wider Adoption: SIEMs are poised for wider adoption beyond large enterprises, becoming a key cybersecurity component for SMEs and critical infrastructure.

END OF Paper 1

Security Information and Event Management (SIEM): Analysis, Trends, and Usage in Critical Infrastructures

2️⃣ Operational Role of Security Information

# The Operational Role of Security Information and Event Management Systems

[Satish Pakalapati](mailto:satishpakalapati65@gmail.com) 21 Sept 2025 08:15 [snp14.pdf](https://drive.google.com/open?id=1gE2jmjb90P-PypHDAffOwnbuifzmL3DB) Completed

Keywords for further research :

Main Use of SIEM Tools

Core Strengths of SIEM

Workflow and Components

Operational Role of SOC Analysts

Challenges of SIEM and SOC Operations

## Introduction

Security Information and Event Management (SIEM) systems represent the backbone of modern cybersecurity operations. They are not standalone defenders but sophisticated tools used by human teams such as Security Operations Centers (SOCs) and Computer Security Incident Response Teams (CSIRTs). While CSIRTs are tasked with responding to security threats and incidents, and SOCs are responsible for monitoring the environment for risks, the SIEM provides the technological foundation.

The SIEM collects, normalizes, stores, and correlates events across the enterprise IT landscape. This enables organizations to detect, analyze, and respond to malicious activities in real time, offering a centralized view of the security posture.

## Use of SIEM Tools

In the early days of cybersecurity, only a handful of systems such as firewalls, intrusion detection systems (IDS), and antivirus software were available for threat assessment. Over time, many more specialized tools emerged, creating two major challenges:

1. Too many interfaces for administrators and analysts to manage.
2. No system capable of correlating events across different tools.

Without correlation, these tools lacked context about the overall system architecture, often triggering false positives. SIEM tools were designed to address these challenges.

They integrate with multiple security systems, collect logs and events, normalize them into a unified schema, and apply rule engines to trigger alerts only when meaningful activity patterns are detected.

## Core Strengths of SIEM

The primary strength of a SIEM lies in its ability to correlate logs from different sources and identify meaningful attack patterns. This makes SIEM act like a radar system for detecting breaches.

Additional strengths include:

* Long-Term Data Retention: Supports forensic investigations and helps detect stealthy threats such as Advanced Persistent Threats (APTs).
* Normalization: Converts raw, vendor-specific logs into a common format for analysis and rule creation.
* Specialized Connectors: Act as translators to parse logs into a consistent format.

## Workflow and Components

Once normalized, data is processed by two key components:

1. Security Management System
   * Applies correlation rules across a sliding event window.
   * Identifies malicious patterns such as brute-force login attempts or port scans.
   * Alerts are displayed on SIEM dashboards for SOC analysts to investigate.
2. Forensic Analysis Database
   * Stores event data for extended periods (from months to years).
   * Supports advanced investigations into slow-moving or hidden attacks.

### Examples of Rule-Based Detection

* Multiple failed login attempts in a short time window
* Continuous requests to access blocked HTTP resources.
* Repeated port-scanning behavior from a suspicious IP address.

Rules may be manually created by analysts or auto-generated by SIEM algorithms using observed patterns.

## Operational Role of SOC Analysts

The SOC analysts (SAs) are the human element working with SIEM outputs. Their work is divided into levels:

* Level I Analysts: Handle large volumes of daily alerts (often 1000–3000). They triage and classify incidents, working in 8–12 hour shifts.
* Level II Analysts: Investigate escalated alerts, refine correlation rules, and identify false positives.
* Level III Analysts: Handle confirmed threats, conduct forensic analysis, and advise on broader defensive strategies.

Because of the workload, Level I analysts often face high burnout rates and typically remain in these roles for only 1–2 years.

## Challenges of SIEM and SOC Operations

Despite their importance, SIEM deployments face several challenges:

1. Contextual Gaps  
    Routine IT tasks like patching or backups may trigger false alerts due to a lack of contextual awareness.
2. Detection of Stealthy Attacks  
    Event monitoring windows may miss slow-moving threats such as APTs. Sampling of raw logs is often necessary to detect such attacks.
3. Scalability  
    Large organizations may generate trillions of events daily. Current infrastructure struggles to achieve the required injection rates for real-time processing.
4. Event Selection  
    With millions of events generated, choosing the right events to log and correlate remains a critical challenge.
5. Storage Requirements  
    Even with data compression, storing trillions of events requires tens of terabytes, making long-term retention costly.
6. Human Resource Costs  
    Running a fully staffed SIEM can cost organizations millions of dollars annually due to technology expenses and analyst requirements.

## Conclusion

SIEM systems are indispensable in providing visibility, correlation, and context across diverse security infrastructures. Their strengths lie in correlation, normalization, and forensic retention, enabling SOCs and CSIRTs to detect and respond effectively to cyber threats.

At the same time, challenges related to scalability, alert fatigue, storage, and contextual awareness highlight the need for continuous innovation. The future direction of SIEM points toward tighter integration with automation platforms (SOAR), greater use of artificial intelligence and machine learning, and enhanced visualization techniques.

In essence, SIEM tools are the radar of enterprise cybersecurity—essential for navigation, but reliant on skilled operators to interpret the signals and steer the organization toward safety.

# END OF Paper 2

# The Operational Role of Security Information and Event Management Systems

3️⃣ Challenges and Directions in Security Information

Challenges and Directions in Security Information and Event Management (SIEM)

[Satish Pakalapati](mailto:satishpakalapati65@gmail.com) [Challenges\_and\_Directions\_in\_Security\_Information\_and\_Event\_Management\_SIEM.pdf](https://drive.google.com/open?id=1CxZIF36T4VGGB7wDVVzsheI8LFprhh3x)

19 Sept 2025 13:30 Completed

Keywords for further research :

II. Reference System → Air Traffic Control SIEM Reference Architecture  
III. Challenges in SIEM → Key Challenges and Limitations of SIEM Systems

IV. Explorative Integration of ATC-SIEM → Practical Integration of SIEM in Air Traffic Control Systems

V. Modeling Behavioral Baselines → Using Machine Learning (LDA) for Behavioral Baselines in SIEM

Abstract

• SIEM is the current standard for handling diverse data sources in cybersecurity.

• The paper explores challenges in using SIEM in a real Air Traffic Control (ATC) system.

• Focuses on issues with massive, unstructured logs and proposes behavioral baselines using Latent Dirichlet Allocation (LDA).

# I. Introduction

• Situational awareness in cybersecurity requires combining diverse data (IDS, logs, audits).

• SIEM helps centralize and normalize this data, forming the backbone of a SOC.

• Current SIEM deployments rely heavily on handcrafted rules and human expertise, which is a limitation.

• Study conducted with LEONARDO’s ATC system, generating very large, unstructured log data.

# II. Reference System

• ATC system provided by LEONARDO for testing under NAPOLI FUTURA project.

• Consists of distributed nodes supporting flight monitoring, conflict detection, and operator stations.

• System produces 15,000+ log lines per minute across ~20 log sources.

• Logs are highly unstructured, verbose, and vary across components.

# III. Challenges in SIEM

• Data Formats: SIEMs work well with structured data, but struggle with unstructured logs (free-text messages).

• Correlation Rules: Analysts must predefine rules, which can’t anticipate unknown or novel attacks.

• Behavioral Baselines: Needed to capture “normal” operations; unstructured logs require NLP/IR approaches.

• Lack of Ground Truth: Incidents are rare, expensive to label, and constantly evolving, making supervised models impractical.

# IV. Explorative Integration of ATC-SIEM

• Used OSSEC, an open-source SIEM/log monitor, as proof-of-concept.

• Developed a connector to wrap proprietary ATC logs into syslog format for SIEM processing.

• Created custom domain rules for detecting ATC-specific patterns.

• Showed feasibility but also limitations: manual rules still dominate, making it hard to detect unknown attacks.

# V. Modeling Behavioral Baselines

• Proposed Latent Dirichlet Allocation (LDA) to automatically model system behavior.

• LDA models topics (system actions) from chunks of log data, without requiring labels or predefined rules.

• Used perplexity to measure how well new logs fit into the learned baseline.

• Experiments showed LDA can distinguish normal vs anomalous chunks, even with low anomaly rates (1–10%).

• Promising because it avoids dependency on human-crafted rules and labeled datasets.

# VI. Conclusion and Future Work

• SIEM in real-world ATC showed both potential and limitations.

• Manual rule-based correlation is insufficient for unstructured, high-volume logs.

• Behavioral baselines with LDA offer an automated way forward.

• Extend SIEMs with plugins (e.g., AlienVault plugins).

• Improve correlation for multi-step attacks.

• Enhance visualization for analysts.

• Develop automation to mimic human analysts’ cognitive processes.

# Acknowledgment

• Supported by STAR program (COSIEM) and RT-CASE project, University of Naples.

END OF PAPER 3

Challenges and Directions in Security Information and Event Management (SIEM)

👾 ML model

Latent Dirichlet Allocation (LDA):

This model is a core focus in Paper 3, which details its use in an Air Traffic Control (ATC) system. The authors propose using LDA to automatically model behavioral baselines from massive, unstructured log data. The model learns "topics" or normal system actions without needing human-labeled data. This is a crucial advancement because it allows the SIEM to detect anomalies (deviations from the baseline) without relying on manually created, static rules.

### What is LDA and Why is it Useful for SIEMs?

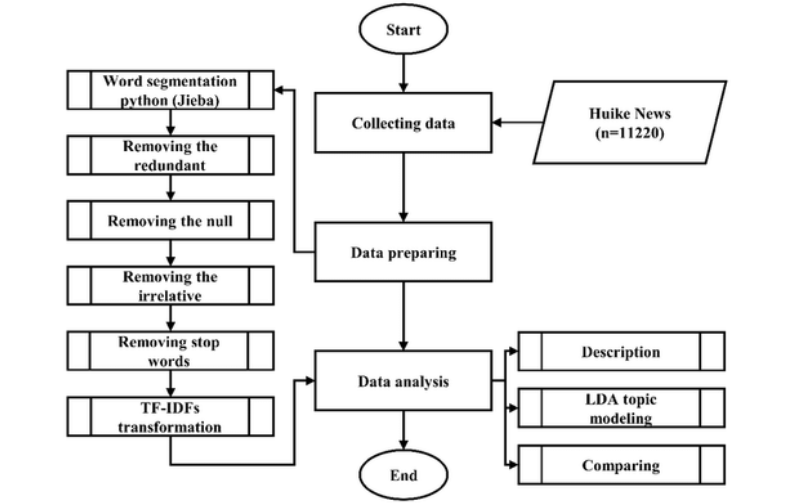
Latent Dirichlet Allocation (LDA) is a statistical, unsupervised machine learning model primarily used for topic modeling in natural language processing (NLP). Its main purpose is to automatically discover the hidden thematic structure (the "topics") within a large collection of documents.

In the context of a SIEM, this concept is applied to cybersecurity logs:

* "Documents" become individual log entries or a collection of log entries from a specific time period.
* "Words" become the discrete terms found within those log entries (e.g., "login," "failed," "user," "admin," "port," "access denied").
* "Topics" become the different types of behavior or activity occurring on the network (e.g., "normal user activity," "database administration," "network scanning").

The key advantage of using an unsupervised model like LDA is that it doesn't require a pre-labeled dataset of "good" and "bad" events. This is perfect for a SIEM, where incidents are rare and expensive to label. Instead, LDA learns what "normal" looks like on its own.

The Working Mechanism of LDA in a SIEM



The process can be broken down into three main phases:

#### 1. Training Phase: Building the Behavioral Baseline

The first phase is the training of the LDA model on a massive corpus of historical, normalized log data. This is how the model builds its understanding of normal network behavior.

* Corpus Creation: The SIEM feeds the LDA model a large collection of log entries, which are treated as "documents." These logs are preprocessed to remove irrelevant data like timestamps and IP addresses, leaving behind only the core descriptive terms.
* Topic Discovery: The LDA algorithm then analyzes the co-occurrence of words within these logs. For example, if the words "login," "success," and "user" frequently appear together, LDA will identify them as belonging to a specific topic, which a human analyst might later label "normal user authentication." Similarly, words like "access denied," "invalid password," and "brute force" might be grouped into another topic, which a human might label "suspicious login attempts."
* Topic Distribution: The output of this phase is a detailed statistical map of the network's behavior. This map includes:
  + A list of all discovered topics.
  + The probability distribution of words for each topic.
  + The probability distribution of topics for each log entry.

This map becomes the behavioral baseline for the network. It's a quantitative representation of what is considered "normal."

2. Detection Phase: Identifying Anomalies

Once the model is trained, it is used to monitor new, incoming log entries in real-time.

* Perplexity Scoring: For every new log entry, the SIEM feeds it to the trained LDA model. The model calculates the entry's perplexity—a measure of how well the new log fits into the existing behavioral baseline.
* Anomaly Flagging: If a new log entry has a very high perplexity score, it means it contains an unusual combination of words and topics that don't fit the learned normal behavior. The SIEM flags this as an anomaly. This could be a novel attack or a user performing an action they have never done before.
* The "Ground Truth" Challenge: This method is powerful because it doesn't need to know what a "bad" event looks like. It only needs to know what a "normal" event looks like. Anything outside of that is considered an anomaly worthy of further investigation.

3. Human Integration: Context and Feedback

The final and most important step is integrating the LDA model's output with the human analyst's expertise.

* Alert Generation: Anomalous log entries are flagged on the SIEM's dashboard with a high risk score. This reduces the number of false positives by filtering out routine activity and highlighting only the statistically significant deviations.
* Contextual Enrichment: Before presenting the alert to the analyst, the SIEM enriches it with data from other sources, such as user identity, asset criticality, and external threat intelligence. This helps the analyst understand the "why" behind the anomaly.
* Feedback Loop: The analyst investigates the alert. If they confirm it as a real threat, the SIEM can use this feedback to refine the model, making it even smarter. If it's a false positive, the analyst can tell the system to adjust its baseline, preventing future false alarms. This continuous feedback loop is what makes the AI-driven SIEM truly adaptive.

4️⃣ A Security pattern

A Security Information and Event Management Pattern

[Satish Pakalapati](mailto:satishpakalapati65@gmail.com) [SIEM\_Pattern.pdf](https://drive.google.com/open?id=1iALDjM1tgH460BED69otHIOWwFOAJ5Ih) 22 Sept 2025 16:30 Completed

Keywords for further research :

SIEM pattern

Security Information and Event Management

Security Operations Center (SOC)

Log collection and normalization

Threat detection and correlation

Security analytics

Forensic investigation

Context enrichment

Alerting and incident response

Automation in cybersecurity

Visual Security Analytics (VSA)

IBM QRadar, Splunk, LogRhythm, McAfee, AlienVault

DINGfest SIEM

Cyber threat intelligence sharing

Modular SIEM architecture

Advanced persistent threats (APT)

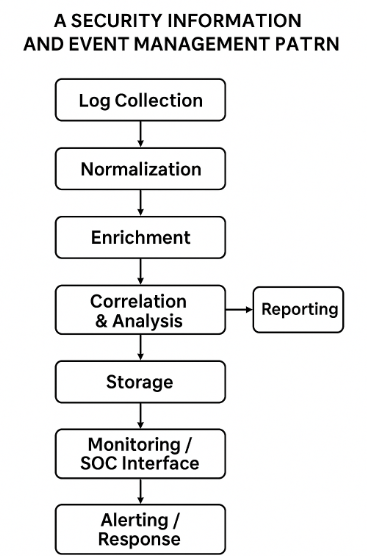
Machine learning in SIEM

### 1. Introduction

* Organizations increasingly face complex IT environments, especially with IoT and Industry 4.0, leading to more attack surfaces.
* Security Information and Event Management (SIEM) systems provide centralized monitoring, enabling security teams to collect, analyze, and respond to incidents.
* Modern SIEM tools are widely used in SOC (Security Operations Centers).
* Problem: Many SIEM solutions exist, but there’s no unified pattern or architecture describing a standard SIEM design.
* Aim of the paper: Propose a SIEM pattern to standardize architecture, improve modularity, reusability, and maintainability.

### 2. Background

* SIEM combines SIM and SEM:
  + SIM (Security Information Management): Historical log collection, compliance, long-term storage.
  + SEM (Security Event Management): Real-time monitoring, threat detection, alerting.
* Modern SIEMs integrate SIM + SEM, offering historical analysis, compliance, alerting, and forensic support.
* Key challenge: Monitoring heterogeneous systems with diverse log formats and high event volumes.



### 3. SIEM Pattern Description

#### 3.1 Intent

* Provide a centralized framework for security event collection, correlation, analysis, and response.
* Ensure automation, human oversight, and integration with multiple security tools.

#### 3.2 Context

* Organizations have complex IT systems: multiple devices, servers, applications, and network layers.
* Standalone security tools (firewalls, antivirus, etc.) are insufficient for advanced threats.
* SOC teams need real-time actionable insights for decision-making.

#### 3.3 Problem / Forces

1. Data Volume & Variety: High volume of heterogeneous logs makes threat detection difficult.
2. Detection Accuracy: Balance between false positives and true threat detection.
3. Reaction Time: Security incidents require fast identification and mitigation.
4. Usability & Integration: SIEM must integrate with new devices easily and require minimal expert effort.
5. Automation: Reduce manual monitoring while enabling automated responses.
6. Analytics & Correlation: Need for rule-based or AI-driven analysis across multiple data sources.
7. Reusability & Modularity: Design must allow reuse of components in different deployments.
8. Cost-effectiveness: Minimize deployment and operational costs.

#### 3.4 Solution / Components

A modular SIEM architecture can include:

1. Log Collection
   * Pull or receive logs from devices using agents or protocols (Syslog, FTP, REST APIs).
   * Handles heterogeneous log formats.
2. Normalization
   * Converts logs into a standardized format.
   * Ensures consistent processing across multiple sources.
3. Enrichment
   * Adds contextual information: geolocation, vulnerability data, user info, threat intelligence.
   * Improves accuracy and prioritization.
4. Correlation & Analysis
   * Detects incidents by correlating events across sources.
   * Can use rule-based methods or machine learning/AI for anomaly detection.
5. Storage
   * Stores both raw and normalized events.
   * Enables historical analysis and forensic investigation.
6. Reporting
   * Generates compliance reports and threat intelligence dashboards.
   * Facilitates management and audit requirements.
7. Monitoring / SOC Interface
   * Provides analysts with visual dashboards and alert management tools.
   * Supports human validation and investigation.
8. Alerting / Response
   * Triggers automated responses or notifies security personnel.
   * May include integration with firewalls, endpoint protection, or ticketing systems.

### 4. Variants / Implementations

* Cloud-based SIEMs: Focus on scalability and real-time processing.
* Some SIEMs change the order of normalization and enrichment depending on use case.
* Examples of commercial SIEMs following this architecture:
  + IBM QRadar
  + Splunk Enterprise Security
  + LogRhythm
  + AlienVault OSSIM / USM

### 5. Consequences

Advantages:

* Improved threat detection and reduced response time.
* Supports automation and SOC efficiency.
* Modular design allows reusability and integration.
* Supports analytics, enrichment, and reporting.

Disadvantages / Limitations:

* Initial implementation effort is high.
* Requires trained personnel for monitoring and analysis.
* Integration of non-standard log sources can be challenging.

### 6. Pattern Mining / Validation

* Steps followed by the authors:
  1. Domain Definition: SIEM systems in IT security.
  2. Information Collection: Whitepapers, manuals, Gartner reports, and SIEM documentation.
  3. Pattern Identification: Use of UML component diagrams to model modular architecture.
  4. Review: Focused on advanced SIEMs for real-world validation.

### 7. Conclusion

* The proposed SIEM pattern standardizes the structure of SIEM systems.
* Benefits include modularity, reusability, maintainability, and automation.
* Future work: Develop detailed patterns for individual modules such as correlation engines or enrichment pipelines.

END OF PAPER 4

5️⃣ Papers Correlation

### How Information Correlates

### 

These four papers are not just a collection of facts; they form a narrative about the life cycle of SIEM technology.

1. What a SIEM Is and Why It Exists: All papers agree that a SIEM is a centralized system for collecting, normalizing, and correlating security logs. The first paper introduces the foundational architecture, and the second and fourth papers expand on this by detailing the specific components like connectors, rule engines, and storage. The core problem they all solve is the overwhelming volume and diversity of security data that makes it impossible for humans to manage on their own.
2. The Human Element (SOCs and CSIRTs): Papers 2, 3, and 4 highlight that a SIEM is only a tool. The real power comes from the Security Operations Center (SOC) team that uses it. Paper 2, in particular, details the tiered structure of SOC analysts (Level 1, 2, and 3) and the high burnout rate they face due to "alert fatigue." This connects directly to the limitations discussed in Paper 1 and 3, which point out that current SIEMs generate too many false positives and rely too heavily on human expertise.
3. Limitations and Challenges: This is a major theme across all papers.
   * Data and Context: Papers 1, 2, and 3 all point out that SIEMs struggle with incomplete data (missing DNS or endpoint logs), unstructured data (free-text logs from industrial systems), and a lack of context that leads to high false positive rates.
   * Rule-Based Flaws: The first, third, and fourth papers all criticize the reliance on basic, manually created rules that cannot detect new or unknown threats like zero-day attacks.
   * Scalability and Storage: Papers 1 and 2 mention the massive scalability challenges of handling trillions of events per day. All papers agree that long-term storage is essential for forensics but is often expensive and difficult to manage.
4. Future Directions and Solutions: The papers offer a unified vision for the future of SIEMs.
   * AI/ML: Papers 1, 3, and 4 all propose that machine learning is the key to overcoming the limitations of manual rules. Paper 3 offers a concrete example of this by using Latent Dirichlet Allocation (LDA) to model system behavior in an Air Traffic Control system.
   * SOAR and Automation: Papers 1, 2, and 4 strongly advocate for integrating SIEMs with SOAR (Security Orchestration, Automation, and Response) platforms to automate repetitive tasks and reduce the human workload.
   * External Factors: Paper 1 provides a unique, high-level PESTLE analysis that shows how external forces—like GDPR regulations, global economic trends, and the rise of IoT—are driving the need for more advanced SIEM capabilities.
5. Usage in Critical Infrastructures: The first and third papers provide a unique application of SIEM technology in Critical Infrastructures (CIs). They highlight the special security needs of these systems (prioritizing availability over confidentiality) and show how SIEMs are being adapted to protect sectors like energy, water, and transportation.

🧑‍🍳 Nikhil

# Research Paper 1 Findings

# [The Operational Role of Security Information and Event Management Systems](https://drive.google.com/file/d/1gE2jmjb90P-PypHDAffOwnbuifzmL3DB/view?usp=sharing)

# [Nikill Reddiee](mailto:bunnynikhilreddy@gmail.com)20 Sept 2025 [snp14.pdf](https://drive.google.com/open?id=1gE2jmjb90P-PypHDAffOwnbuifzmL3DB) Completed

**Security Information and Event Management (SIEM) systems** within a **Security Operations Center (SOC)**. A SOC is a centralized unit that monitors security events in real time to protect an enterprise's IT assets. The main job of a SIEM system is to collect, normalize, and analyze security events from many different sources, like firewalls and application servers. By correlating events from these diverse sources, the system can identify potential malicious activity and trigger alerts for human analysts to review.

**Structuring SOCs around SIEM Systems:**

* The system receives inputs from various security devices and sensors, such as network firewalls, intrusion prevention systems, and host sensors.
* These devices output security events with unusual or anomalous behavior that may indicate malicious intent.
* Because these events are in vendor-, device-, and version-specific formats , the SIEM system's first task is to normalize them into a common format for easier processing and rule creation.
* SIEM system connectors, customized for each device type and vendor, receive and convert the events into a common format in a scalable manner.

Once the events are normalized, they are forwarded to two main components:

* **The security management platform:** This platform analyzes a rolling window of events, typically those from the past few hours. Its rule engine periodically applies rules to these events. A triggered rule generates a new alert, which is sent to the SIEM system terminal for a SOC analyst to review. Rules can be created by an analyst or algorithmically generated by the system, for example, through pattern mining or anomaly detection.
* **The archival forensic analysis database:** This database stores events for a longer term, such as three to six months, for forensic investigation. A typical archival database can ingest up to 100 Kbytes per second, whereas a typical management platform can ingest around 10 Kbytes per second.

### **Technical Challenges and Opportunities**

SIEM systems face significant technical challenges, primarily due to the massive scale of events generated by modern enterprise networks. The number of events is proportional to the number of employees, devices, and applications, and this volume is expected to keep growing.

* **Event Collection**: It's a logistical challenge to collect events at the necessary scale, with a target ingestion rate of up to 12 million events per second for a network generating 1 trillion events per day. SIEM systems also struggle with noisy data and the risk of attackers injecting malicious events. A key question remains: how to determine the "right" events to collect without fearing you're missing something important?
* **Event Storage**: SIEM systems have to balance high storage costs with analysis requirements. Storing data indefinitely is impractical, so a data retention period is necessary. There is also a trade-off in storage architecture, as the needs for real-time analysis (quick reads) and forensic analysis (high ingestion rates) are different.
* **Event Correlation and Analysis**: Identifying attacks from event streams is considered "more art than science". Analysis algorithms must continuously evolve to keep up with changing threats and employee behaviors. The problem of false positives is made worse by the scale of data, as benign events outnumber malicious ones by orders of magnitude. Even true positives can overwhelm analysts, so SIEM systems need to prioritize alerts.
* **Visualization**: Since SIEM systems won't replace human analysts, they need to improve their ability to summarize analysis results and present them effectively. This includes developing visualization techniques that help humans process large amounts of data and provide timely context.

### **Opportunities for Improvement**

* **Leveraging Computer Science Advances**: The document suggests that advances in fields like storage systems (e.g., nonvolatile memory) and distributed computing, especially for big data analysis, will significantly impact SIEM systems. This could lead to more scalable analysis and more complex correlation rules.
* **Enhancing Alert Correlation and Prioritization**: Although alert correlation has been a long-standing area of research, the large amount of low-quality data can lead to false correlations. This has prompted a focus on alert prioritization to help analysts concentrate on higher-quality alerts.
* **Creating an Adaptive, Holistic Solution**: The paper concludes that SIEM systems need to become more adaptive and context-aware. Future solutions should:
  + Allow SOCs to adapt their processes for each incident.
  + Centralize incident-related information to help the system learn from past events and reduce repetitive investigations.
  + Automate repetitive tasks by providing analysts with contextual information alongside alerts.

# Research Paper 2 Findings

**Challenges and Directions in Security Information and Event Management (SIEM)**

# [Nikill Reddiee](mailto:bunnynikhilreddy@gmail.com)21 Sept 2025 Completed

I. INTRODUCTION

SIEM (Security Information and Event Management) is a crucial component of a Security Operations Center (SOC). It helps to collect and normalize various data sources for security analysis. However, integrating SIEM is challenging.

II. REFERENCE SYSTEM

* The ATC system was developed as part of the "Novel Approaches to Protect Critical Infrastructures from Cyber Attacks" (NAPOLI FUTURA) project, an academia-industry collaboration.
* The system is composed of several distributed nodes (e.g., D02, MN1, and SFN) that perform critical ATC functions like flight progress monitoring and conflict detection.
* It operates in a controlled testing environment and is exercised with test suites that simulate typical ATC operations such as flight creation, re-routing, and removal.
* The system generates a massive volume of proprietary text logs for various purposes, including control-flow tracing and event reporting.
* There are approximately 20 different log sources across the system's 7 nodes, including both standard syslogs and legacy application logs.
* Application logs do not have a mandated structured format. An example from the D02-PAN log source shows lines with timestamps, messages, and other data.

III. CHALLENGES IN SIEM:

**Data Formats**

* Modern SIEM frameworks use internal formats like MDI Fabric and Common Information Model to import and combine data from many sources.
* While these frameworks handle structured data well with built-in adapters, their capabilities are limited for unstructured text. For example, in syslog data, they only automatically recognize a few standard fields like timestamp and hostname, leaving the message itself in its raw format.

**Correlation Rules**

* SIEM products allow analysts to set and monitor data correlation rules for detecting compromises. However, analysts must define these rules themselves.
* This traditional approach requires humans to know "in advance" what they're looking for.
* The challenge is that security attacks often create patterns across different data sources that are unknown and difficult for analysts to envision beforehand

IV. EXPLORATIVE INTEGRATION OF ATC-SIEM:

**Tool of Choice**: The researchers chose OSSEC, an open-source tool that is also a component of the well-known OSSIM open-source SIEM. OSSEC has built-in alerting for standard data sources like syslog, Apache..

**Creating a Connector**: To handle the ATC system's proprietary, unstructured application logs, a custom connector program was developed. This program:

* Polls for new lines in a log file at a specific sampling rate.
* Wraps each new log line into a standard syslog message format that OSSEC can accept.
* Forwards the wrapped message to the OSSEC server.

**Leveraging Existing Capabilities**: By converting the proprietary logs into syslog format, the researchers were able to leverage OSSEC's built-in syslog monitoring and alerting features. They extended the  
 local\_rules.xml file to add new, custom correlation rules.

**Proof-of-Concept Rule**: A proof-of-concept SIEM rule was created to trigger an alert whenever a log line contains the keyword "calcnextsecgeog". The corresponding alert shows the syslog-wrapped message alongside the original log line.

**Dual Detection**: This integration allows OSSEC to detect not only domain-specific compromises based on the custom rules but also conventional compromises tracked by standard log sources, such as syslogs, that the ATC system already emits

V. CONCLUSION AND FUTURE WORK:.

**Using plugins:** They will explore using plugins to extend the capabilities of traditional SIEM systems, noting that platforms like AlienVault support a large number of existing plugins and the creation of new ones.

**Expanding experiments:** They plan to validate their initial findings and address correlation techniques for detecting stealthy and multi-step attacks across logs.

**Visualization:** This is another area of future research they will investigate.

**Developing automatons:** They plan to create proof-of-concept automatons to mimic the cognitive work of security analysts when investigating data.

# Research Paper 3 Findings

**A Security Information and Event Management Pattern**

# [Nikill Reddiee](mailto:bunnynikhilreddy@gmail.com)22 Sept 2025 [SIEM\_Pattern.pdf](https://drive.google.com/open?id=1iALDjM1tgH460BED69otHIOWwFOAJ5Ih) Completed

**1. INTRODUCTION**

* Protecting corporate IT infrastructure from cyberattacks is becoming more difficult due to trends like Industry 4.0 and the Internet of Things, which increase the number of attack points.
* Most mid-to-large-sized companies have a **Security Operations Center (SOC)** to get a complete, centralized view of IT security and react quickly to incidents.
* Over 80% of SOCs are supported by a **SIEM (Security Information and Event Management) system** to improve IT security awareness.
* SIEM systems automate incident detection and response by collecting security data at a central location to provide a holistic view of an organization's IT security. They also enable historical and correlated analysis and other security measures.
* The authors of this paper propose a pattern to define the generic structure of a SIEM system, aiming to fill the gap of a missing standardized pattern in the market.
* The goal of this pattern is to help developers define generic modules, improving the reusability, replaceability, and overall structure of SIEM software.
* The pattern was developed as part of the DINGfest project, which focuses on next-generation SIEM systems, including collecting forensic evidence from virtual environments, leveraging Identity and Access Management data, and improving Visual Security Analytics.

2. SECURITY INFORMATION AND EVENT MANAGEMENT

**SIM vs. SEM**: SIM focuses on the centralized collection, storage, and management of historical log data for compliance and reporting. SEM, on the other hand, deals with real-time threat management, monitoring security incidents, and triggering immediate reactions.

**Modern SIEM**: Today's SIEM has evolved to be a more holistic and integrated solution that combines the advantages of both SIM and SEM.

**Core Functionalities**: SIEM systems collect relevant security data from a wide variety of event and contextual data sources. They perform historical analyses, support compliance reporting, and assist in forensic investigations. Their main functions are their broad scope of event sources and their ability to correlate and analyze events from these heterogeneous sources.

**Purpose**: The central purpose of collecting and analyzing a large amount of data is to identify anomalies and incidents that would be invisible if data from only a single system were considered. It also facilitates the monitoring and management of an organization's overall IT security state.

**Integration with SOCs**: A SIEM system is typically integrated into a Security Operations Center (SOC). As described by Bhatt et al., a SOC is a centralized unit that monitors all security-related events for an IT infrastructure.

**Role of the SOC**: A SOC consists of software (such as a SIEM system), processes, and a team of security analysts. The SOC team determines the appropriate actions to take after a potential incident is detected. SOCs are often organized in hierarchical layers and may work in multiple shifts to provide around-the-clock incident response.

**Automation**: As the volume of log data and the need for trained security analysts grow, there is an increasing demand for a higher level of automation. SIEM systems can help by automatically analyzing and reacting to incidents in a centralized way. They also assist analysts by better visualizing large amounts of data to leverage their expertise more efficiently.

3. SIEM PATTERN

SIEM, SIEM system, (Big Data) Security Analytics System, Cyber Threat Intelligence Tool/System

The proposed SIEM pattern consists of eight main components that work together in a structured way:

* **Log Collection**: Gather security data from various sources. This can be done by "pulling" logs from sources or by having the sources "push" logs to the system.
* **Enrichment**: Adds extra context to the raw log data. For example, it might translate an IP address into a geolocation to make it more useful for analysts.
* **Normalization**: Converts all the different log formats from various devices into a single, standardized format to make it easier for subsequent components to process.
* **Correlation & Analysis**: This is the central part where the system identifies incidents by analyzing the incoming normalized data in real-time and also performs historical analysis for forensic purposes. It can use various methods, including rule-based detection.
* **Storage**: Stores both the raw log data and the analysis results.
* **Reporting**: Generates reports for internal use or to meet external regulatory and compliance requirements.
* **Monitoring**: Provides a user interface for human analysts to review and interact with the data and alerts.
* **Alerting & Incident Response**: Automatically triggers reactions or sends alerts to the appropriate people in case of a detected incident.

### **Forces and Challenges the Pattern Addresses**

The pattern is designed to address several key forces and challenges:

* **Detection Rate and False Positives**: The system aims for a high detection rate and a very low false-positive rate to reduce the manual workload on security analysts.
* **Speed of Detection and Reaction**: It enables fast detection and reaction to attacks to minimize financial damage.
* **Usability**: The system should be easy for analysts to use, with simple processes for connecting new devices and creating new rules.
* **Automation**: It seeks to automate analysis and incident response to reduce the number of staff required, as well-trained security experts are rare.
* **Holistic View**: The pattern allows for the connection of a large number of diverse devices to provide a comprehensive view of an organization's security.
* **Integration of Expert Knowledge**: The monitoring module provides a way for human experts to analyze data, customize automatic analyses, and provide input.

### **Examples of the Pattern in Use**

The paper shows that the proposed pattern aligns with the architecture of several popular SIEM systems:

* **IBM QRadar**: QRadar's architecture, with its layers of "data collection," "data processing," and "data searches," fits the proposed pattern, although it uses different, brand-specific names for its components.
* **DINGfest Project**: A research project for a next-generation SIEM system, DINGfest, also largely fits the pattern but includes unique features like **pseudonymization** for privacy and the ability to collect snapshots of virtual machines for forensics.

### **Advantages and Disadvantages**

**Advantages**:

* **Improved Detection**: The pattern's correlated analysis and use of context data can lead to a higher detection rate and a lower false-positive rate.
* **Faster Response**: Centralizing security events makes it easier to react quickly, and the pattern allows for automated incident response.
* **Cost Reduction**: Automation and a single monitoring interface can help reduce staff and overall costs.
* **Reusability**: The modular design with defined interfaces promotes reusability and intelligence sharing between different systems and organizations.

**Disadvantages**:

* **High Introduction Effort**: Integrating a SIEM system can be complex and demanding, especially in organizations with historically grown IT infrastructures.
* **Need for Experts**: Despite automation, there is still a need for security experts to set up and monitor the system.
* **Limited Human Input**: The pattern focuses on machine-generated data and doesn't provide a way for humans to directly input valuable information they might have observed (e.g., a suspicious phone call).

4. PATTERN MINING / IDENTIFICATION PROCESS

**Domain Definition**: The initial step was to define the domain of the pattern, which is **SIEM**.

**Coverage Consideration**: The scope was narrowed down to focus only on the **software components** of SIEM, excluding surrounding processes handled by a SOC.

**Information Format Design**: The pattern was designed using a **UML component diagram**, chosen because it fits the desired level of abstraction and is well-known in software engineering.

**Information Collection**: A variety of sources were used to gather information, including:

* **Commercial SIEM products**: This was considered the most important source to ensure the pattern reflects the current state-of-the-art.
* **Research papers**: Used to incorporate the current academic view on SIEM.
* **Whitepapers, manuals, and documentations**: These provided detailed information on product functionalities and architecture.
* **Product websites**: Used as an additional source, though with caution, as they are often marketing-focused.

**Information Review**: To refine the information, the authors reviewed leading SIEM systems identified in the **Gartner Magic Quadrant** for SIEM. The analyzed systems included IBM QRadar, Splunk Enterprise Security, LogRhythm Enterprise, HPE ArcSight, and McAfee Enterprise Security Manager. Additionally, the open-source SIEM **OSSIM** was reviewed to gain deeper insight

5. CONCLUSION AND FUTURE WORK

* **Pattern Creation**: The authors deduced a generic SIEM pattern by following a formal pattern identification process. They analyzed advanced SIEM systems on the market and identified the forces they depend on.
* **Pattern Details**: The pattern is visualized as a UML component diagram, and its advantages and disadvantages are discussed. To provide more insight, the pattern was compared to a commercial SIEM system and a research project's SIEM.
* **Application**: The paper is intended to help SIEM tool developers by providing a structured way to define modules, which can improve the overall system architecture. It can also serve as a foundation for those new to the field to understand SIEM's functional principles and assembly.
* **Future Work**: The proposed pattern is not detailed enough on its own. Future research is necessary to create more patterns that explain each module in more detail and break them down into sub-modules.

# Research Paper 4 Findings

**Security Information and Event Management (SIEM): Analysis, Trends, and Usage in Critical Infrastructures**

# [Nikill Reddiee](mailto:bunnynikhilreddy@gmail.com)23 Sept 2025 In Progress

**1. Introduction**

Due to increased activity from nation-states and cybercriminals, cybersecurity risks have grown significantly. The timely detection of these sophisticated threats has become a major challenge for industrial control systems (ICT) and IT infrastructure.

**Current Threats**: Common cybersecurity incidents include ransomware, malware that disrupts business operations, phishing campaigns targeting privileged users, business email compromises, data theft, and social engineering.

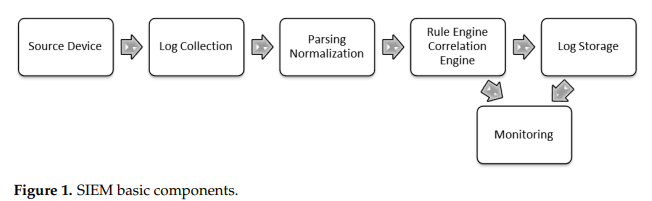
**SIEM Capabilities**: According to a NIST report, cybersecurity solutions for industrial control systems should offer real-time behavioral anomaly detection, faster incident management, and intelligent network visualization. Security Information and Event Management (SIEM) systems have these capabilities as built-in features.

**Function of SIEM**: SIEMs collect, store, and correlate events from multiple sensors, such as intrusion detection systems, antivirus software, and firewalls. They serve as the central platform for modern security operations centers (SOCs), providing synthetic views of alerts for threat handling and reporting.

**Market Landscape**: The SIEM market includes products from a variety of companies, including traditional IT companies (e.g., HP, IBM), those with more visionary approaches (e.g., AT&T Cybersecurity/AlienVault), and promising tools like Splunk.

**2. SIEM Solutions**

**Security Information and Event Management (SIEM) systems** help administrators design security policies and manage events from various sources. A basic SIEM is made up of several parts, including log collection, normalization, a rule engine, and log storage, all of which must work together to function properly. SIEM platforms provide real-time analysis of security events from network devices and applications. While newer SIEMs can automate some responses, they often don't perform a complete analysis of an attack's impact before taking action.



### **SIEM Classification and Market Trends**

SIEM vendors are classified by well-known institutions like

**Gartner** in its annual **Magic Quadrant** report. Gartner categorizes vendors into four groups:

* **Leaders**: Companies that perform well and are well-positioned for the future.
* **Challengers**: Organizations that perform well now but might not fully understand the market's future direction.
* **Visionaries**: Organizations that have a good vision for the market but don't yet execute their plans well.
* **Niche Players**: Organizations that focus on a small market segment or are unfocused and don't outperform competitors.

Other organizations, like Techtarget and Info-Tech Research Group, also analyze SIEMs, often using Gartner's Magic Quadrant as a basis for their assessments.

Over the last decade, companies like **IBM** (QRadar), **McAfee**, **LogRhythm**, and **RSA** have consistently been at the top of Gartner's list. Some companies have merged to adapt to market changes, such as the merger of HP and ArcSight. Recently, SIEM solutions have begun to include

**User and Entity Behavior Analytics (UEBA)** and smart dashboards as innovations to help security administrators with reports, advanced analytics, and incident response.

### **Strengths and Weaknesses of Major SIEM Tools**

* **ArcSight Enterprise Security Manager**: Provides a graphical interface for SOC teams but has limited visualization and complex correlation rules.
* **QRadar (IBM)**: A versatile tool that can be deployed in multiple ways (hardware, software, virtual, SaaS). It offers a user-friendly interface and supports threat intelligence feeds but relies on third-party technologies for advanced features like endpoint monitoring.
* **McAfee Enterprise Security Manager**: Offers a scalable architecture with real-time forensics and a risk-based correlation engine. However, its predictive analytics and built-in behavioral analysis are not well-developed and require additional solutions.
* **LogRhythm NextGEN SIEM Platform**: Provides endpoint monitoring and UEBA capabilities. It is not ideal for critical infrastructures without extensions and requires a high degree of automation.
* **USM and OSSIM (AT&T Cybersecurity/AlienVault)**: Offers both commercial and open-source options with a web-based interface. It has limited UEBA and machine learning capabilities, and its reaction features are basic and restricted to predefined conditions.
* **RSA Netwitness Platform**: Focuses on advanced threat detection by analyzing network data, including logs and packets. It has strong operational technology (OT) monitoring capabilities but requires a deep understanding of its options due to its complexity and cost.
* **Splunk Enterprise Security**: A market leader in operational intelligence with advanced security analytics and machine learning. Its predefined correlation rules are basic, and it relies on third-party applications for task automation.
* **SolarWinds Log and Event Manager**: Provides centralized log collection, real-time correlation, and intuitive visualization. It lacks support for public cloud services and custom report writing.

🌹 Prem

# The Operational Role of Security Information and Event Management Systems

[Prem Swaroop](mailto:premswaroop129@gmail.com)22 Sept 2025 08:00[snp14.pdf](https://drive.google.com/file/d/1kHbb8au7rEcg7eeu6wsyQcPtj1SZGG5j/view?usp=sharing)Completed

# **Key Words :**

Security Information and Event Management (SIEM)

Security Operations Center (SOC)

Computer Security Incident Response Team (CSIRT)

Intrusion Detection Systems (IDS)

Firewalls

Advanced Persistent Threats (APTs)

Event Correlation

**Abstract :**

The paper discusses the operational role of **SIEM** systems in enterprise-level Security Operations Centers (**SOCs**) and **CSIRT**s. SIEM systems are vital for monitoring security events across networks, servers, applications, and user accounts. They collect, normalize, store, and correlate logs from multiple sources to detect malicious activities. The paper highlights the strengths of SIEMs (cross-correlation, long-term storage for forensics, alert detection) and the challenges (scalability, false positives, lack of contextual information, storage and analysis bottlenecks). It concludes that while SIEM systems are indispensable, further advances in big data analytics, visualization, storage, and automation are required for them to effectively support large enterprise security needs.

**Operations Performed in the Research Paper (SIEM Systems) :**

**1. Event Collection**

SIEM systems gather security-related logs and events from different sources such as firewalls, IDS/IPS, servers, authentication systems, DNS, and proxies. This ensures that all activities across the network are monitored in one place.

**2. Normalization**

Since logs come in various vendor-specific formats, SIEM converts them into a standardized schema. This makes further processing, analysis, and correlation easier and more efficient.

**3. Event Storage**

Normalized events are stored securely for two main purposes: real-time monitoring of ongoing incidents and long-term storage for forensic investigation.

**4. Event Correlation**

SIEM correlates multiple events using rule-based engines to identify malicious behavior patterns. These rules may be predefined by analysts or automatically generated through techniques like anomaly detection and pattern mining.

**5. Alert Management**

When suspicious activities are detected, SIEM generates alerts for Security Operations Center (SOC) analysts. Analysts investigate, classify, and escalate cases while refining detection rules to reduce false positives.

**6. Forensic & Big Data Analysis**

SIEM helps in investigating past incidents by analyzing large amounts of stored data. It also detects advanced persistent threats (APTs) that move slowly over time and identifies long-term attack trends.

**7. Visualization & Decision Support**

Finally, SIEM provides dashboards, graphs, and reports that present correlated insights. This visualization helps SOC analysts make quicker and more accurate security decisions.

**Tools and Technologies Mentioned in the Research Paper :**

**SIEM Systems:** Security Information and Event Management platforms collect, normalize, store, and correlate security events from multiple sources to provide centralized monitoring and detection.

**Firewalls & Intrusion Prevention Systems (IPS/IDS):** Firewalls block unauthorized access while IDS/IPS monitor network traffic to detect and prevent malicious activity.

**Antivirus Systems (AVS):** Antivirus software identifies, blocks, and removes malware or malicious programs from endpoints and servers.

**Authentication Systems & Web Application Firewalls:** Authentication systems verify user identities, while WAFs protect web applications from threats like SQL injection, cross-site scripting, and unauthorized access.

**DNS and Proxy Logs:** Logs from DNS and proxy servers provide insights into user web activity, helping detect suspicious domains, phishing attempts, or unusual traffic patterns.

**Big Data Analytics:** Techniques like clustering, pattern mining, and anomaly detection are applied on large-scale event data to uncover hidden attack patterns and advanced threats.

**Visualization Techniques:** Dashboards, charts, and graphs present complex security data in an easy-to-understand form for SOC analysts, improving incident response.

**Archival Forensic Databases:** These databases store historical log data for long-term retention, allowing investigators to perform deep forensic analysis after an attack.

**Challenges Highlighted :**

SIEM systems face several significant challenges in managing enterprise security. One major issue is **scalability**, as these systems must process **trillions of events daily**, which demands enormous **storage** and **processing capabilities**. Another concern is **false positives**, where the high volume of **alerts** can overwhelm analysts and reduce overall **efficiency**. SIEMs also suffer from a lack of context, since important internal activities such as **patching** or **maintenance** are often missing, leading to potential **misinterpretation of events**.

Additionally, there is **underuse of long-term data**; stored logs are rarely leveraged for detecting **advanced persistent threats (APTs)**. Challenges in event collection and validation arise when noisy or **maliciously injected data** complicates analysis. Organizations also face **storage trade-offs**, balancing the need for **long-term forensic retention** against high storage costs. Finally, human factors pose a challenge, as **SOC analysts** often experience burnout, high turnover, and heavy manual workloads due to continuous alert triage.

**Conclusion :**

The paper concludes that SIEM systems are the “radar” of cybersecurity, essential for detecting attacks, correlating logs, and guiding forensic analysis. However, without addressing scalability, false positives, contextual awareness, and automation, SOCs risk being overwhelmed. The authors call for advances in big data, AI/ML, and visualization to make SIEM systems more adaptive and sustainable in enterprise environments.

**END OF THE PAPER**

# **Challenges and Directions in Security Information and Event Management (SIEM)**

[Prem Swaroop](mailto:premswaroop129@gmail.com)22 Sept 2025 20:30 [Challenges\_and\_Directions\_in\_Security\_Information\_and\_Event\_Management\_SIEM.pdf](https://drive.google.com/file/d/1-nukNYYCr4Tbx83j6glADeNV7rI_vxZQ/view?usp=sharing)Completed

**Keywords :**

Security Information and Event Management (SIEM)

Air Traffic Control (ATC) Systems

Behavioral Baselines

Latent Dirichlet Allocation (LDA)

**Abstract :**

This research paper explores the challenges and future directions of Security Information and Event Management (SIEM) systems in handling heterogeneous and unstructured log data. Using a real-world Air Traffic Control (ATC) system as a case study, the authors highlight the complexity of managing massive, unstructured logs and integrating them into SIEM frameworks. They present the use of open-source tools like OSSEC for log monitoring, and propose modeling behavioral baselines using Latent Dirichlet Allocation (LDA) to detect anomalies without relying solely on pre-defined correlation rules. This work suggests that data discovery approaches and automated baselines can significantly improve SIEM effectiveness.

**Introduction :**

Cybersecurity relies on collecting, processing, and correlating diverse data sources to build situational awareness. SIEM systems are the backbone of Security Operations Centers (SOC), enabling detection and response to threats. However, real-world deployment faces challenges:

Data variety and unstructured logs make normalization difficult.

Correlation rules depend on expert knowledge and struggle with unknown attack patterns.

Human analysts remain central, as automation is still limited.

The paper focuses on an ATC system developed by LEONARDO, which generates over 15,000 log lines per minute. By analyzing this case, the authors highlight SIEM limitations and propose new directions for anomaly detection.

**Key Challenges in SIEM :**

SIEM revolves around **Data Formats**, where handling and normalizing unstructured logs is far more difficult than working with structured data. Another major issue lies in **Correlation Rules**, since analysts are required to anticipate attack patterns in advance, which is often unrealistic and limits detection of unknown threats. A further challenge is the development of **Behavioral Baselines**, as it is necessary to model normal system behavior instead of relying solely on pre-defined rules. Finally, the **Lack of Ground Truth** makes it difficult to train supervised models, as labeled incident data are rarely available in real-world environments.

**Tools Mentioned in this Research Paper :**

**OSSEC:** An open-source Host-based Intrusion Detection System (HIDS) for log monitoring and alerting. In this study, it was adapted to handle ATC logs by converting them into syslog format and adding custom rules.

**OSSIM:** An open-source SIEM platform by AlienVault, integrating tools like OSSEC and Snort. It provides centralized event correlation and threat detection.

**Apache Storm:** A real-time stream processing system used here to analyze massive ATC log data (15,000+ lines per minute) efficiently.

**Latent Dirichlet Allocation (LDA):** A topic modeling technique applied to logs for building behavioral baselines. It helps detect anomalies as deviations from normal system behavior.

**R Packages** **:**

topicmodels: For fitting LDA models.

text2vec: For text processing and log data preparation.

LDAvis: For interactive visualization of topics.

**Important Contributions :**

The paper makes several key contributions to advancing SIEM practices. It first demonstrates the integration of OSSEC into an Air Traffic Control (ATC) system, enabling domain-specific monitoring of unstructured logs. It then proposes behavioral baseline modeling using Latent Dirichlet Allocation (LDA), allowing anomaly detection without relying on predefined rules. The study further validates that LDA-based detection is highly sensitive, even at low anomaly levels, ensuring accurate identification of irregular behavior. Finally, it suggests that data discovery combined with automation can effectively complement traditional SIEM approaches, reducing dependency on manual correlation rules and expert knowledge.

**Conclusion :**

This concludes that current SIEM practices require enhancement through automated, data-driven approaches. Future research could focus on extending SIEM capabilities with plugins, improving the detection of stealthy and multi-step attacks, and enhancing visualization techniques for analysts. Additionally, developing automated cognitive agents to replicate human expertise in log analysis presents a promising direction for advancing security monitoring and incident response.

**END OF THE PAPER**

# **Security Information and Event Management (SIEM) Pattern**

[Prem Swaroop](mailto:premswaroop129@gmail.com)23 Sept 2025 10:30[A%20Security%20Information%20and%20Event%20Management%20Pattern.pdf](https://drive.google.com/file/d/1gPcywLGdAYLDCjZ31Ut_KzBC6tTrlyQp/view?usp=sharing)Completed

**Keywords :**

Security Information and Event Management (SIEM)

Security Operations Center (SOC)

Log / Event

Pattern - Oriented Software Architecture (POSA)

Threat Intelligence

### **Introduction :**

Protecting corporate IT infrastructures from cyber threats has become increasingly difficult due to growing complexity from trends like the Internet of Things (IoT). To manage this, most medium to large companies establish a**Security Operations Center (SOC)**, with over 80% of them using a **SIEM system**.

The core purpose of a SIEM is to create a single, centralized view of an organization's IT security. It achieves this by collecting security data from many different systems, which allows for the detection of advanced threats that might otherwise go unnoticed. SIEM systems automate threat detection, improve reaction times, and help analysts manage the vast amount of security data generated daily. Despite the widespread use of SIEMs, there was no generic pattern defining their fundamental structure. This paper aims to fill that gap by proposing a reusable pattern for developers and a foundational guide for newcomers to the field.

### **Main Topics Covered in the Research Paper :**

**General Overview of SIEM**: The paper explains that SIEM evolved from two separate concepts: Security Information Management (SIM), for log storage and compliance reporting, and Security Event Management (SEM), for real-time threat monitoring. Modern SIEMs integrate both functions.

**The Proposed SIEM Pattern**: The central contribution of the paper is a detailed, eight-component pattern that describes the architecture of a generic SIEM system.

**Pattern Implementation and Examples**: The authors provide practical implementation details for each component of the pattern. They also compare their generic pattern to two real-world examples: the commercial product  
 **IBM QRadar** and the research project **DINGfest**.

**Pattern Mining Process**: The paper outlines the bottom-up methodology used to identify and create the pattern, which involved analyzing market-leading products, research papers, and technical documentation.

### **Core Problem and Purpose :**

Modern corporate IT environments are incredibly complex and have a growing number of potential attack points. To manage this, companies use a**Security Operations Center (SOC)**, which relies heavily on a SIEM system to get a centralized view of security. A SIEM's job is to collect and analyze security data from many different sources to detect advanced threats and enable quick responses to incidents.

Despite many SIEM tools on the market, the authors found there was no standardized, generic pattern that defined the essential structure of a SIEM. This paper's main goal is to create such a pattern by analyzing existing systems and research. This pattern helps define the core software modules, which can increase reusability and improve the overall structure of SIEM systems.

### **The Proposed SIEM Pattern :**

The central contribution of the paper is an eight-component pattern that outlines the flow of data and key functions within a typical SIEM system.

Here is a breakdown of each component:

* **Log Collection**: This is the starting point, where the system gathers security-relevant data (logs and events) from a wide variety of sources like firewalls, servers, and applications.
* **Enrichment**: Raw log data is often cryptic. This module adds valuable context to make it more understandable and useful. For example, it could translate an IP address into a real-world location or add user role information to an authentication log.
* **Normalization**: Logs come in many different formats. The normalization component transforms all incoming data into a single, standardized format. This is critical because it allows the analysis engine to use a single set of rules for all data, regardless of its origin.
* **Correlation & Analysis**: This is the "brain" of the SIEM. It analyzes the processed data in real-time to identify threats and incidents. It correlates events across multiple systems to spot complex attacks that would be invisible if looking at a single system's logs. It also performs historical analysis for forensic purposes.
* **Storage**: This component stores all relevant data, including raw logs, normalized data, and the results of analyses. This historical data is crucial for compliance reporting and investigating past incidents.
* **Reporting**: This module creates reports and allows for the sharing of **threat intelligence** with other systems or organizations. It also helps generate reports required for regulatory compliance.
* **Monitoring**: This provides the user interface for human analysts. Here, experts can view the prepared data, investigate potential threats, fine-tune detection rules, and identify false positives.
* **Alerting & Incident Response**: When an **incident** (a security breach or violation) is confirmed, this component triggers alerts to notify the correct personnel. It can also initiate automated actions to help contain the threat.

### **Validation and Methodology :**

To prove their pattern is accurate and reflects real-world systems, the authors compared it to two examples:

1. **IBM QRadar**: A leading commercial SIEM product. The paper shows that QRadar's architecture, while using different terminology, aligns with all the major components of the proposed pattern.
2. **DINGfest**: A next-generation SIEM system from a research project. DINGfest also fits the pattern but adds unique features like data pseudonymization for privacy and advanced visual analytics.

The authors developed this pattern using a **bottom-up identification process**. They systematically reviewed market-leading SIEM products (like Splunk and McAfee), scientific research papers, product manuals, and whitepapers to identify the common, recurring components that form the state-of-the-art in SIEM architecture.

**END OF THE PAPER**

# 

# 

# 

**SIEM Analysis, Trends and Usage**

[**Prem Swaroop**](mailto:premswaroop129@gmail.com)**23 Sept 2025 05:00**[**sensors-21-04759-with-cover.pdf**](https://drive.google.com/file/d/1ShOdW4cV9DbIRju-TkEOJDmt_f-DoJNz/view?usp=sharing)**Completed**

## **Keywords :**

## Evolution of SIEMs

## SIEM enhancement

# SIEM trends

# Critical infrastructures

## **Introduction :**

This document summarizes the research paper "Security Information and Event Management (SIEM): Analysis, Trends, and Usage in Critical Infrastructures".**Security Information and Event Management (SIEM)** systems are powerful tools used to prevent, detect, and respond to cyber-attacks by collecting, correlating, and analyzing security event data from various sources. Cybersecurity risks have grown significantly, making timely and sophisticated detection a major challenge. This paper provides a comprehensive analysis of the most widely used SIEM solutions, evaluates their features and limitations, examines external factors influencing their evolution, and proposes enhancements for the next generation of SIEMs, particularly for use in critical infrastructures.

## **Core Concepts of the Research Paper :**

### **1. SIEM Solutions and Classification**

A SIEM system is built from several core components that work together, including a source device, log collection, parsing and normalization, a rule/correlation engine, log storage, and monitoring. The paper analyzes SIEM solutions by categorizing them according to the

**Gartner Magic Quadrant**, which classifies vendors as:

* **Leaders**: Organizations that are well-positioned for the future and execute well on their current vision.
* **Challengers**: Organizations that execute well today but may not demonstrate a clear understanding of market direction.
* **Visionaries**: Organizations that understand where the market is headed but do not yet execute well.
* **Niche Players**: Organizations that focus successfully on a small market segment.

Key long-standing vendors mentioned include

**IBM (Qradar)**, **RSA (NetWitness)**, **MicroFocus (ArcSight)**, and **LogRhythm**.

### **2. Key SIEM Features and Capabilities**

The paper evaluates SIEM tools based on a set of critical features, including:

* **Correlation Rules**: The ability to write complex rules to detect security incidents.
* **Data Sources**: The capacity to natively collect events from diverse sources like firewalls, servers, and antivirus systems.
* **Real-Time Processing**: The capability to handle and analyze millions of events in real-time to prevent or react to incidents.
* **Visualization**: The tools provided for creating custom dashboards and interactive exploration of security data.
* **Data Analytics**: The use of anomaly detection and machine learning to analyze the behavior of users and applications.
* **Reaction and Reporting**: The native ability of a SIEM to trigger actions (e.g., send emails, run scripts) in response to security incidents.

### **3. Limitations of Current SIEM Systems**

Despite their power, current SIEMs have several significant limitations:

* **Incomplete Data**: It is not cost-effective to capture all data, leading to fragmented information that can result in false positives and negatives.
* **Basic Correlation Rules**: Most SIEMs use weak, basic boolean logic that is insufficient for detecting sophisticated, multi-stage attacks.
* **Limited Storage**: Archived data is often difficult to access and use, and the archiving process is frequently manual and costly.
* **Reliance on Humans**: Many threat responses still require slow, expensive, and error-prone manual analysis by human experts.
* **Basic Reaction Capabilities**: Automated responses are often limited to simple actions like sending an email or opening a ticket, requiring add-ons for more complex reactions.
* **Limited Visualization**: Reporting and visualization capabilities are often too generic and rudimentary to provide deep, actionable insights from the massive amount of collected data.

### **4. The Future of SIEMs: A PESTLE Analysis**

The paper analyzes external factors that will shape the future of SIEMs:

* **Political**: Increased government investment in cybersecurity will empower the adoption and evolution of SIEM technology.
* **Economic**: The rise of freelancers and SMEs will drive demand for user-friendly, service-based SIEM models.
* **Societal**: Society's growing dependency on technology makes SIEMs essential for protecting against cyber-attacks on critical services.
* **Technological**: Advances in **cloud storage**, **big data analytics**, and **machine learning** are key enablers that will make SIEMs smarter and more powerful.
* **Legal**: Data privacy regulations like **GDPR** require SIEMs to handle personal data in a compliant and protected manner.
* **Environmental**: The explosion of **IoT devices** and cloud services will drastically increase the volume of data that SIEMs must collect and analyze.

### **5. Proposed Enhancements for Future SIEMs**

To address current limitations, the next generation of SIEMs should focus on these enhancements:

* **Integration with SOAR**: Combine SIEM with **Security Orchestration, Automation, and Response (SOAR)** platforms to automate complex incident response workflows.
* **AI/ML Capabilities**: Integrate **Artificial Intelligence and Machine Learning** to provide predictive analytics, improve the detection of abnormal behavior, and reduce false positives.
* **Enhanced Visualization**: Develop flexible and interactive visualization models capable of handling diverse, real-time data streams to support better decision-making.
* **Enhanced Storage**: Leverage secure and elastic cloud solutions for long-term data archival, allowing for customizable data retention policies.
* **OSINT Data Fusion**: Incorporate **Open-Source Intelligence (OSINT)** by using language processing to identify threats discussed on social media and the deep web.

### **6. SIEM Usage in Critical Infrastructures**

Protecting critical infrastructures (CIs) is a unique challenge because their systems prioritize

**availability** (ensuring the system is always running) over the confidentiality and integrity emphasized in traditional IT. SIEMs are essential for monitoring and protecting CIs, but they often require manual intervention due to the sensitive nature of these environments. The paper details the use of SIEMs in several sectors:

* **Energy**: Detecting attacks like GPS spoofing and anomalies in power grid signal strengths.
* **Water Supply**: Monitoring SCADA networks in real-time to detect deviations in water quality that could indicate an attack.
* **Transportation**: Protecting digitized transport networks from data breaches and disruptions to physical assets.
* **Healthcare**: Providing a consolidated view of security events to prevent data breaches and ensure compliance with regulations like HIPAA.
* **Financial Services**: Detecting insider threats, data exfiltration, and account abuse using User and Entity Behavior Analytics (UEBA).

**THE END OF THE PAPER**

🪙 Abijith

# **The Operational Role of Security Information and Event Management Systems**

# 

# [Abijith Chowdary Billa](mailto:abijithchowdary@gmail.com) Sep 22, 2025 11:00 AM Completed

# 

# SIEM - Security Information and Event Management (Tool)

# SOCs - Security Operations Center (Place or a Center)

# CSIRTs - Computer Security Incident Response Teams (Team)

# 

# In modern cybersecurity, CSIRTs are the teams responsible for handling security incidents, while SOCs serve as the central hub that actively monitors an organization's IT assets, such as firewalls, servers, and user accounts, for any signs of danger. The primary tool that empowers these centers is the SIEM system, which works by gathering, standardizing, storing, and connecting security events to pinpoint malicious activities.

# 

# The Core Function of SIEM Tools:

# Initially, organizations relied on a handful of separate security tools like firewalls and antivirus software, each with its own interface. This led to two major problems:

# 

# It became overwhelming to manage so many different dashboards.

# 

# These tools operated in isolation, with no ability to share information or connect events across the system.

# 

# This isolation meant that individual tools, unaware of the broader IT architecture, would often generate false alarms. SIEM systems were developed specifically to solve this. They act as a central brain, taking in data from all these different security tools, converting it into a common language, and using correlation rules to identify real threats from the noise.

# 

# The key strength of a SIEM is its ability to cross-reference logs from various sources to identify complex attack patterns. When a suspicious sequence of events occurs, it alerts the Security Analysts (SAs). Essentially, a SIEM acts like a radar for the network. Furthermore, its capacity to store events for long periods is invaluable for forensic investigations and for detecting slow, stealthy attacks like Advanced Persistent Threats (APTs).

# 

# How SIEMs Process Information:

# Since SIEMs collect data in various formats from numerous sources, their first crucial step is normalization—converting all the data into a standard format. This is done by Specialized Connectors that parse the incoming logs.

# 

# Once normalized, the data is sent to two places:

# 

# The Security Management Platform for immediate analysis.

# 

# The Forensic Analysis Database for long-term storage.

# 

# The management platform typically analyzes a short window of recent events (a few hours), while the database can archive data for months. The database can handle a much higher data ingestion rate (around 100 KBps) compared to the management platform (around 10 KBps). The platform's rule engine constantly checks these recent events against defined rules. When a rule is matched, an alert is sent to the SOC analysts.

# 

# Rules are designed to spot malicious behavior, such as:

# 

# A burst of failed login attempts in a short time.

# 

# HTTP requests to known malicious websites.

# 

# Network scanning activity.

# 

# These rules can be created manually by analysts or generated automatically by the SIEM itself by spotting frequent event patterns.

# 

# The Human Element: SOC Analysts:

# SOC analysts work in tiers. Level 1 SAs are the first line of defense, monitoring the SIEM console and triaging between 1,000 to 3,000 alerts per day. Due to the high-pressure environment and long shifts, job retention for Level 1 analysts is often short. Their main task is to decide if an alert is a real threat or a false alarm. Uncertain cases are escalated to Level 2 SAs, who perform a deeper investigation using additional internal and external sources. If a real breach is confirmed, they escalate it to higher-level forensics and engineering teams.

# 

# The Cost and Challenges of SIEM Operations:

# For large global enterprises, running a SIEM is a major investment. The initial hardware setup can cost $3-5 million, with annual operating costs for a team of 20 analysts adding another $5 million or more. This significant investment is justified by the critical protection SIEMs provide, but they also come with operational challenges.

# 

# Key Challenges Faced by SOCs Using SIEMs:

# 

# Rule Management and False Alarms: Writing precise rules is a constant battle. Overly specific rules might miss attacks, while generic rules flood analysts with false positives, creating a vicious cycle of alert overload.

# 

# Lack of Contextual Information: SOCs are often isolated from routine IT operations. This means that normal activities like system patching can trigger security alerts, forcing analysts to waste time tracking down false alarms because they lack automated context about network changes.

# 

# Detecting Stealthy Attacks: SOCs primarily focus on real-time alerts from a narrow time window. This makes it very difficult to detect slow, advanced attacks (APTs) that unfold over months. Some teams try to mitigate this by randomly sampling old logs, but a more systematic approach is needed.

# 

# Scalability and Data Management: The volume of events is enormous. To collect one trillion events per day, a SIEM would need to ingest 12 million events per second—a rate far beyond current capabilities. A major challenge is figuring out which events are the "right" ones to collect without missing crucial evidence.

# 

# Storage Limitations: Storing all this data is expensive. For example, one trillion events could require 40 Terabytes of storage daily even with compression. This forces a tough trade-off between storage costs and the need to retain data for future analysis.

# 

# 

# 

# **Challenges and Directions in Security Information and Event Management (SIEM)**

# [**Abijith Chowdary Billa**](mailto:abijithchowdary@gmail.com)Sep 22, 2025 7:00 PM **Completed**

# **SIEM** – Security Information and Event Management **SOC** – Security Operations Center **LDA** – Latent Dirichlet Allocation **ATC** – Air Traffic Control

# This paper shares a real-world industry experience of integrating a SIEM into a critical **Air Traffic Control (ATC)** system developed by Leonardo, a leading aerospace and defense company. The system generates massive amounts of highly unstructured text logs, making security monitoring particularly challenging.

# **Key Challenges in SIEM Deployment:**

# **Unstructured Logs:** The logs from the ATC system are verbose, non-standard, and lack a consistent format. This makes it difficult to parse and analyze them using traditional SIEM tools without significant customization.

# **Lack of Ground Truth:** Unlike many IT systems, there is no labeled dataset (ground truth) of security incidents for ATC systems. This makes it impossible to use supervised machine learning techniques to detect anomalies or attacks.

# **No Mature Threat Model:** The ATC domain lacks a well-established threat model. This means security analysts cannot rely on predefined attack patterns or correlation rules, which are commonly used in traditional SIEM deployments.

# **Human Dependency:** Setting up effective SIEM rules requires deep domain knowledge and continuous human input. Analysts must manually define what constitutes “suspicious activity,” which is time-consuming and error-prone.

# **Exploratory Integration with OSSEC:**

# The authors used **OSSEC**, an open-source host-based intrusion detection and log monitoring tool, to integrate SIEM capabilities into the ATC system. Since OSSEC did not natively support the proprietary log format, a custom **connector** was developed to:

# · Poll log files at a fixed rate.

# Wrap each log line into a standard syslog format.

# Forward the logs to the OSSEC server for analysis.

# A sample rule was created to trigger an alert when the keyword "calcnextsecgeog" appeared in the logs. While this demonstrated basic monitoring capability, it highlighted the limitation of relying solely on predefined rules.

# **A Novel Approach: Behavioral Baseline Using LDA:**

# To overcome the limitations of rule-based detection, the authors proposed an unsupervised learning approach using **Latent Dirichlet Allocation (LDA)**, a topic modeling technique. Here’s how it worked:

# **Logs as Documents:** Logs were split into 10-second chunks, each treated as a “document.”

# **Topic Modeling:** LDA was used to identify recurring “topics” (i.e., patterns of system behavior) from these log chunks.

# **Perplexity as an Anomaly Score:** The model was trained on normal operational logs. During detection, if a new log chunk had a high **perplexity** value (meaning it didn’t fit the learned topics), it was flagged as anomalous.

# This method required **no labeled data**, **no prior threat knowledge**, and was **fully automated**. It effectively identified deviations from normal behavior without relying on hardcoded rules.

# **Experimental Results:**

# The optimal number of topics (K) for the ATC logs was found to be 12.

# Anomalous logs—injected with unknown keywords—were correctly identified with high perplexity scores.

# The approach showed promise in detecting unknown attacks or unusual system behaviors that rule-based SIEMs would miss.

# **Conclusion and Future Work:**

# The study concludes that while SIEMs are essential for security monitoring, they must evolve beyond manual rule-setting. Data-driven, unsupervised methods like LDA can **complement traditional SIEMs** by automatically modeling normal behavior and flagging deviations.

# Future directions include:

# Developing SIEM plugins for automated anomaly detection.

# Extending the approach to detect multi-step or stealthy attacks.

# Improving visualization tools to help analysts interpret results.

# Building cognitive systems that mimic human analytical reasoning.

# This work highlights the need for **adaptive, intelligent SIEM systems**—especially in critical infrastructures where threats are evolving and log data is complex and unstructured.

# 

# 

# 

# 

# 

# 

# 

# 

# **A Security Information and Event Management Pattern**

# [Abijith Chowdary Billa](mailto:abijithchowdary@gmail.com) Sep 23, 2025 4:30 PM Completed

# SIEM – Security Information and Event Management

# SOC – Security Operations Center

# SIM – Security Information Management

# SEM – Security Event Management

# LDA – Latent Dirichlet Allocation

# This paper introduces a generic architectural pattern for SIEM systems, derived by analyzing existing market solutions and research initiatives. The goal is to provide a reusable blueprint that helps developers and organizations understand, design, and implement SIEM systems in a structured and modular way.

# Introduction to SIEM

# SIEM systems are central to modern Security Operations Centers (SOCs). They collect, normalize, and analyze security-related data from various sources to detect threats and support incident response. The term SIEM combines:

# SIM (Security Information Management): Focuses on long-term log storage, reporting, and compliance.

# SEM (Security Event Management): Deals with real-time monitoring, correlation, and alerting.

# Modern SIEM systems integrate both functionalities into a unified platform.

# The SIEM Pattern

# The authors propose a generic SIEM pattern structured into eight core components, visualized using a UML component diagram:

# Log Collection:

# Gathers logs from diverse sources such as firewalls, servers, and applications. Data can be collected via push (e.g., syslog) or pull (e.g., SCP, FTP) methods.

# Enrichment:

# Enhances raw log data with contextual information (e.g., geolocation of IP addresses, asset details, vulnerability data) to improve analysis.

# Normalization:

# Converts heterogeneous log formats into a standardized structure for consistent processing.

# Correlation & Analysis:

# The core component where events are analyzed in real-time and historically using rules, machine learning, or statistical methods to detect incidents.

# Storage:

# Archives normalized logs, analysis results, and historical data for forensic and compliance purposes.

# Reporting:

# Generates compliance reports and shares threat intelligence with external systems (e.g., STIX format).

# Monitoring:

# Provides a user interface for security analysts to visualize data, investigate alerts, and fine-tune detection rules.

# Alerting & Incident Response:

# Automatically notifies stakeholders or triggers responses when threats are detected.

# Example Implementations

# IBM QRadar

# Fits the pattern closely with layers for data collection, processing, and search/interface.

# Uses event collectors, flow collectors, risk and vulnerability managers, and a custom rules engine.

# DINGfest (Research SIEM)

# A next-gen SIEM with a focus on virtual machine introspection, identity behavior analytics, and privacy via pseudonymization.

# Lacks built-in reporting and automated response but extends traditional SIEM capabilities with advanced analytics and forensic evidence collection.

# Forces and Trade-Offs

# The pattern addresses several conflicting requirements:

# Detection Rate vs. False Positives: Better correlation improves detection but must minimize false alarms.

# Automation vs. Human Oversight: Reduces manual effort but still requires expert intervention.

# Scalability vs. Cost: Must handle high data volumes affordably.

# Usability vs. Complexity: Should be easy to use despite underlying complexity.

# Advantages of the Pattern

# Improved Detection: Cross-source correlation and context enrichment enhance threat visibility.

# Faster Response: Centralized analysis enables quicker incident handling.

# Modularity: Components can be replaced or upgraded independently.

# Compliance Support: Built-in reporting for regulatory requirements.

# Reusability: Standardized interfaces support integration and sharing.

# Liabilities

# High Implementation Effort: Complex and heterogeneous IT environments make deployment challenging.

# Dependence on Experts: Still requires skilled analysts for tuning and oversight.

# Limited Human Input: Cannot incorporate informal observations (e.g., phishing phone calls).

# Related Patterns

# IDS Pattern: Focuses on monitoring network traffic rather than log analysis.

# Centralized Logging Pattern: Collects logs but does not perform security analysis.

# Adapter Pattern: Useful for integrating incompatible data sources or modules.

# Pattern Identification Process

# The authors used a structured approach:

# Domain Definition: SIEM and SOC context.

# Coverage Limitation: Focus on software architecture, not processes.

# Information Format: UML component diagrams.

# Information Collection: Analyzed market leaders (IBM QRadar, Splunk, LogRhythm, etc.) and research projects.

# Information Review: Based on Gartner’s Magic Quadrant and open-source tools like OSSIM.

# Conclusion and Future Work

# This paper provides a high-level, reusable pattern for SIEM systems that reflects current best practices. It helps in understanding SIEM architecture and supports modular development. Future work will involve defining more detailed sub-patterns for each component and incorporating emerging trends like visual analytics and cloud-native SIEM architectures.

# 

# 

# 

# 

# **Security Information and Event Management (SIEM): Analysis, Trends, and Usage in Critical Infrastructures**

# [**Abijith Chowdary Billa**](mailto:abijithchowdary@gmail.com)Sep 23, 2025 9:00 PM **Completed**

# **Authors:**

# Gustavo González-Granadillo, Susana González-Zarzosa, Rodrigo Diaz

# **Journal:**

# Sensors (2021)

# 

# **Introduction**

# SIEM systems are critical for preventing, detecting, and responding to cyber-attacks.

# They provide visibility into high-risk areas and support proactive mitigation strategies.

# SIEMs are evolving to integrate with big data analytics and AI/ML technologies.

# 

# **SIEM Solutions**

# **SIEM Classification (Gartner Magic Quadrant)**

# Leaders: IBM QRadar, Splunk, ArcSight, LogRhythm

# Challengers: McAfee, RSA

# Visionaries: AlienVault/AT&T Cybersecurity, Exabeam

# Niche Players: SolarWinds, Fortinet, ManageEngine

# 

# **SIEM Tools Overview**

# ArcSight: Strong correlation, complex setup

# QRadar: Good threat intelligence, basic reaction capabilities

# McAfee ESM: Scalable, requires add-ons for advanced features

# LogRhythm: UEBA and forensics, limited critical infrastructure support

# 

# AlienVault USM/OSSIM: Open-source option, limited ML capabilitie

# RSA NetWitness: Advanced threat detection, complex configuration

# Splunk: Strong analytics and visualization, expensive

# SolarWinds: User-friendly, limited cloud support

# 

# **SIEM Features and Capabilities**

# **Key functionalities evaluated:**

# Correlation rules

# Data sources support

# Real-time processing

# Data volume handling

# Visualization

# Data analytics

# Performance

# Forensics

# Complexity

# Scalability

# Risk analysis

# Storage

# Price

# Resilience

# Reaction and reporting

# UEBA

# Security automation

# 

# **Limitations of Current SIEMs**

# **Incomplete Data**

# SIEMs lack full visibility due to cost and privacy constraints (e.g., GDPR).

# Identity fragmentation and shared accounts reduce correlation accuracy.

# 

# **Basic Correlation Rules**

# Most SIEMs use simple Boolean logic; few support advanced historical correlation.

# 

# **Basic Storage Capabilities**

# Limited archival options; cloud storage is underutilized due to security concerns.

# 

# **Reliance on Humans**

# High dependency on expert analysts for threat response.

# 

# **Basic Reaction and Reporting**

# Limited automated response; lack of standardized incident reporting.

# 

# **Limited Data Visualization**

# Generic dashboards; poor support for real-time, multi-source data visualization.

# 

# **The Future of SIEMs (PESTLE Analysis)**

# **Political**

# Increased EU investment in cybersecurity (€1.8B by 2020) supports SIEM adoption.

# 

# **Economic**

# Growth in freelance work and cybersecurity jobs (35% by 2020) drives demand for user-friendly SIEMs.

# 

# **Societal**

# Generation Z’s tech awareness and social media growth influence SIEM data sources.

# 

# **Technological**

# Cloud, IoT, 5G, AI/ML, and big data analytics will shape next-gen SIEMs.

# 

# **Legal**

# GDPR compliance affects data processing and storage in SIEMs.

# 

# **Environmental**

# IoT and cloud adoption increase data volume and attack surfaces.

# 

# **Potential Enhancements for Future SIEMs**

# **Diverse Security**

# Use of diverse defenses and probabilistic risk metrics.

# 

# **OSINT Data Fusion**

# Integrate open-source threat intelligence using NLP.

# 

# **Enhanced Visualization**

# Interactive, real-time dashboards for multi-dimensional data.

# 

# **Enhanced Storage**

# Secure, elastic cloud archival solutions.

# 

# **Integration with SOAR**

# Automate response workflows with Security Orchestration, Automation, and Response.

# 

# **AI/ML Capabilities**

# Improve detection accuracy and reduce false positives.

# 

# **Other Enhancements**

# Risk-based metrics, hierarchical SIEM architectures, XDR integration, and autonomous operation.

# 

# **SIEMs in Critical Infrastructures**

# **Energy Distribution**

# Detects attacks like DDoS, GPS spoofing, and sleep deprivation.

# 

# **Water Supply**

# Monitors SCADA systems for water quality changes and malicious activity.

# 

# **Transportation**

# Protects digitized transport networks from cyber-physical threats.

# 

# **Healthcare**

# Ensures HIPAA compliance and protects electronic health records.

# 

# **Financial Services**

# Detects insider threats, fraud, and supports regulatory reporting.

# 

# **Related Work**

# Gartner, Techtarget, and Info-Tech provide commercial evaluations but lack technical depth.

# Academic works focus on specific enhancements (e.g., UEBA, attack modeling) but not holistic SIEM evolution.

# 

# **Conclusions**

# SIEMs are evolving into data analytics platforms with AI/ML integration.

# Enhancements needed in visualization, storage, automation, and critical infrastructure support.

# Future SIEMs must be scalable, compliant, and user-friendly to meet growing cybersecurity demands.

# 

# 

# 

🌪️ Kamesh

**The Operational Role of Security Information and Event Management Systems**

1. CSIRTs( “Computer security incident response teams” ) are dedicated teams that handle cybersecurity incidents.

2. Every CSIRT serves a specific area or organization.

3. The CSIRT continuously watches the organization’s systems, networks, and applications to spot suspicious activities early.

4. The SOC( “Security Operation Center” ) is the command center where analysts sit and monitor the organization’s security. It supports the CSIRT by doing day-to-day monitoring and alerting.

5. SOCs use SIEM systems to collect, normalize and analyze logs/events from many sources (firewalls, servers, apps). SIEMs help detect attacks faster.

But as organizations grow, SIEMs must handle more data and bigger infrastructures (scalability challenge).

**Security Incident and Event Management Systems**

A Security Operations Center (SOC) acts as the command center for an organization’s security, monitoring all parts of the IT environment, including networks, firewalls, intrusion prevention systems, servers, databases, and user accounts. The SOC collects logs and event data from these systems and triggers alerts whenever unusual activity is detected.

SOC analysts review each alert to determine whether it is a false alarm—such as routine maintenance—or a real threat. False alarms are ignored or closed, while genuine threats are escalated to the incident response team. In severe cases, involving data theft or critical breaches, SOCs may also engage HR, legal, public relations, or law enforcement.

To operate effectively, a SOC requires strong analytics and forensic tools, real-time threat intelligence, deep knowledge of its own network and systems, and clear internal processes to respond quickly to incidents.

Before SIEM (Security Information and Event Management) systems, each security tool had its own dashboard, making it difficult for analysts to correlate alerts across multiple systems. Single tools often produced many false positives, whereas correlating alerts from multiple tools increases the likelihood of detecting real attacks.

A SIEM system centralizes security monitoring by collecting logs from all tools, normalizing them into a common format, and applying correlation rules to detect meaningful patterns. SIEMs also maintain an inventory of network assets to rank alerts by importance and provide long-term log retention for forensic analysis and detection of stealthy attacks, such as advanced persistent threats (APTs). Its main strength lies in correlating events across systems, effectively “scanning the sky” for threats.

**Structuring SOCs around SIEM Systems**

A SIEM (Security Information and Event Management) system collects data, including logs and events, from various security tools such as perimeter defenses (firewalls, intrusion prevention), host sensors (antivirus, intrusion detection), applications (web app firewalls, authentication systems), and network sensors. Each tool produces events in its own format, so the SIEM first normalizes these into a common format, making them easier to process and apply rules.

SIEM uses connectors—special plugins or modules—for each type of tool or vendor. These connectors read, parse, and normalize events, and the system must be scalable to handle high volumes of data quickly. Once normalized, events are managed in two main ways: the management platform handles short-term monitoring and alerting (few hours), while the archival database stores long-term logs (months or years) for forensics and compliance purposes. Archival systems can handle much larger data volumes, and specialized hardware can further improve performance.

The SIEM’s rule engine evaluates incoming events against predefined patterns of suspicious or malicious activity. When a rule matches, an alert is generated for SOC analysts. Common rules include detecting too many failed logins (possible brute-force attacks) or HTTP requests to known malicious sites (malware callbacks).

Rules in SIEM can be manual (written by analysts), automatic (learned from patterns using machine learning), or anomaly-based (triggered when activity deviates from a normal baseline). This combination of collection, normalization, and intelligent alerting allows SIEMs to provide comprehensive, real-time visibility into security events across an organization.

**SOC Structure**

A Security Operations Center (SOC) is built around a SIEM system and staffed by Security Analysts (SAs) organized into levels based on experience. Level 1 analysts are entry-level, Level 2 are more experienced, and Level 3 and above include specialists in forensics and security engineering.

Level 1 SAs constantly monitor the SIEM alert dashboard, determining whether alerts are real threats or false positives. If unsure or overwhelmed by false alarms, they escalate issues to higher levels. These analysts handle large volumes of alerts—often 1,000–3,000 per day—and work in 24×7 shifts, leading to high stress and turnover, with many leaving within two years.

Level 2 SAs perform deeper investigations, leveraging additional internal and external data sources. They escalate confirmed breaches to Level 3 or forensics teams, and for false alarms, they assist engineers in fine-tuning SIEM rules to reduce noise and improve detection accuracy.

**Cost-Benefit Analysis**

The popularity of SIEM systems in SOCs is driven by their ability to handle huge volumes of events from diverse sources. In the past, when enterprise networks were smaller with fewer devices, applications, and users, SIEM was less necessary. As networks grew, the number of events from firewalls, IDS, applications, servers, and other sources exploded, making SIEM indispensable over the last decade.

Running a SIEM at enterprise scale is expensive. A typical management platform can cost around $80,000, while an archival database costs about $20,000. Large deployments often require hundreds of connectors, multiple platforms, and numerous archival databases, with upfront hardware costs reaching $3–$5 million. On top of that, yearly maintenance and staffing costs—such as 20 security analysts—can push operating expenses over $5 million annually.

Because companies invest millions to operate a SIEM and SOC, these systems must deliver tangible value. They need to provide timely detection, reduce false positives, and enable fast incident response. Without these benefits, management may question the justification for such high costs.

**Operational Challenges**

Security Operations Centers (SOCs) face challenges with SIEM tools due to the large scale of enterprise networks, the complexity of diverse devices and log formats, and the high rate of incoming events from sensors, firewalls, and servers, making it difficult to ingest, process, correlate, and respond to all alerts effectively.

**Rule Creation and Management:** SOC analysts gain better visibility with access to all logs, but the huge volume of data creates big challenges in storage, search, sharing, transfer, and analysis. SIEM struggles to run complex searches quickly, and even small false-positive rates generate many alerts, causing alert fatigue. Analysts must balance specific rules (catch one attack) versus generic rules (catch more but create false alarms), and limited staff often slows triage, making it hard to keep up with explosive data growth.

**Lack of Contextual Information:** SOC teams don’t handle routine IT tasks, but SIEM still generates alerts for them, creating unnecessary work. Having up-to-date info on devices, servers, and users could automatically identify routine changes and reduce false positives. Currently, communication is often manual (verbal or email), causing inefficiency and lost context during shift changes or staff turnover. Automating the storage and sharing of operational info from IT/network teams to SOC would improve efficiency, reduce false alarms, speed up triage, and enhance threat detection.

**Ad Hoc Use of Long-Term Data:** SOCs typically focus on recent events, ignoring SIEM’s long-term log storage, which makes detecting slow attacks like Advanced Persistent Threats (APTs) difficult. Relying on a rolling window or manual sampling of old logs is inefficient and unreliable. Analysts should use long-term logs with analytics and visualization tools to uncover hidden trends and unusual patterns over weeks or months. This approach helps identify new attack patterns, improves SIEM rules, and allows SOCs to shift from reactive alert handling to proactive threat detection.

**Technical Challenges and Opportunities**

SIEM remains central to SOCs, but as networks grow larger and more complex, it must improve its ability to collect, store, analyze, and visualize massive amounts of event data. Event logs capture device activity, software processes, and user behavior, and their volume is enormous due to many devices, complex applications, numerous employees, and multiple events generated per user action. Larger organizations, IoT devices, BYOD policies, and new hires further increase log volume, creating exponential growth in the data SOCs must manage.

**Event Collection:** Scaling log collection for SIEM is extremely challenging because modern networks can generate millions of events per second, requiring numerous connectors, servers, bandwidth, and storage, making it very expensive. Some sources provide too little data, while others produce excessive noise, reducing SIEM effectiveness. Logs also vary in format, may have missing or incorrect fields, or could include maliciously injected events, so SIEM must validate integrity, verify sources, and flag suspicious data. Collecting everything is costly, but collecting too little risks missing attacks. SOCs must balance coverage by prioritizing critical events, possibly sampling less important logs, while ensuring they don’t miss key indicators of compromise.

**Event Storage:** Keeping all logs indefinitely is ideal for forensics but extremely expensive, as modern networks can generate terabytes to petabytes of data each month. SIEM must define a retention period that balances cost, usefulness, and legal requirements like GDPR, supporting automatic deletion or anonymization. Forensic analysis needs fast ingestion but can tolerate slower reads, while real-time monitoring requires fast reads for dashboards and alerts. Logs often contain sensitive data (IP addresses, usernames, URLs), so SIEM must encrypt them, control access, and audit usage to prevent breaches.

**Event Correlation and Analysis:** SIEM faces major challenges in large enterprises, including correlating multiple event streams at scale, deriving true value from massive data through sophisticated analysis, and detecting attacks where no fixed definition exists. Continuous learning is required as adversaries adapt and networks evolve. High volumes of benign events lead to false positives, while even perfect detection can overwhelm analysts, making alert prioritization essential. Spurious correlations in high-dimensional data create misleading alerts, and the hardest challenge is inferring attacker intent, as malicious and benign activities can produce similar patterns.

**Visualization:** SIEM systems cannot fully replace human analysts, who remain critical for judgment and decision-making in SOCs. SIEM acts as a support tool, helping identify new attacks and prioritize alerts. It must present analysis clearly, provide timely context, and use advanced visualizations to reveal patterns and anomalies in large datasets. Outputs should be tailored for multiple audiences, from system administrators needing operational detail to management requiring strategic summaries, enabling humans to act efficiently and effectively.

**Toward Addressing the Challenges**

SIEM vendors offer improvements in data collection, storage, and correlation, but scalability remains a major challenge, and complex correlation rules are often avoided, allowing sophisticated or stealthy attacks to go undetected. Direct research on SIEM is limited, though advances in storage systems, parallel and distributed computing, and distributed correlation engines offer potential solutions. Research approaches include big data clustering to detect anomalies, combining multiple events for alert correlation, and using probability-based methods to prioritize alerts. Big data and security-focused visualization techniques are also being developed, helping SOC analysts interpret SIEM data more effectively.

**A Security Information and Event Management Pattern**

**Introduction**

With the rise of Industry 4.0 and IoT, protecting corporate IT systems has become more complex, leading many organizations to establish Security Operations Centers (SOCs) that rely heavily on SIEM (Security Information and Event Management) systems. SIEM centralizes security data, automates incident detection, supports rapid responses, and enables historical and correlated analysis for deeper insights. Despite many products, there is no standard SIEM structure; this paper proposes a SIEM pattern to identify core components, improve modularity, and enhance compatibility. As part of the DINGfest project, it focuses on next-generation SIEM architecture, forensic evidence collection, IAM data integration, and improved visual analytics, using the POSA template to extract best practices for developers, administrators, and beginners with basic IT security knowledge.

**Security Information and Event Management**

SIEM (Security Information and Event Management), first introduced by Gartner, combines the strengths of SIM (Security Information Management) — centralized collection and preservation of historical logs for compliance and audit reporting — and SEM (Security Event Management) — real-time monitoring, threat detection, and incident response. Using data aggregation to reduce volume, SIEM evolved into a holistic, centralized platform that collects data from diverse sources, performs historical analysis, supports compliance reporting and forensics, and correlates events across heterogeneous systems to detect anomalies and monitor an organization’s overall security posture.

Typically deployed within a Security Operations Center (SOC), which is a centralized unit combining software, processes, and skilled analysts, SIEM enables log collection and analysis from all assets, determines post-incident actions, and supports hierarchical, multi-shift operations for 24/7 monitoring. As log sources and security complexity grow, SIEM increasingly relies on automation to analyze and react to incidents while visualizing large data volumes to enhance analyst expertise.

**SIEM Pattern**

SIEM systems, also known as Security Information and Event Management (SIEM), (Big Data) Security Analytics Systems, or Cyber Threat Intelligence Tools/Systems, are designed to centrally collect all security-relevant data to detect threats and incidents. They handle diverse data formats from various security devices, enable both real-time and historical analysis of security events, and provide interaction with human security experts for informed decision-making. In modern organizations with numerous evolving security systems and devices, isolated tools like firewalls or antivirus software only detect specific attacks, so a centralized SIEM is essential to maintain a holistic view of IT security. Real-time analysis enables quick responses to ongoing threats, while historical analysis helps identify patterns and investigate past incidents.

**A ) Problem**

Organizations face several challenges in managing IT security due to the sheer number of devices and software systems, as well as the large volumes of log data. Key issues include:

1. Complexity of detecting sophisticated attacks:
   * Advanced attacks, often called Advanced Persistent Threats (APTs), can affect multiple systems simultaneously and may go undetected.
2. Data overload:
   * Massive log data complicates analysis and response, making it difficult to react promptly to incidents.

**Forces/Challenges :**

* **Detection rate and false positives:** Automatic detection should reduce analyst workload, achieve high threat coverage, and minimize false alerts that waste expert time.
* **Time of discovery and reaction:** Fast detection and response are vital to limit damage, with systems automating reactions or promptly notifying the right personnel.
* **Usability:** The system must be easy to use, support new device integration and rule creation, and require minimal expert knowledge for daily operations.
* **Visual preparation of data & expert knowledge integration:** Visual analytics and context enrichment should integrate human expertise to detect novel, previously unseen incidents.
* **Degree of automation:** Because skilled security staff are scarce and costly, analysis and responses should be automated as much as possible.
* **Number of connectible devices:** The system should connect all relevant devices, handle many log formats, and manage large data volumes for a holistic view.
* **Analytics capabilities:** It must support rule creation, event correlation, and both manual and automated (human-assisted) analyses to uncover complex threats.
* **Reusability and intelligence sharing:** The system should be modular for easy integration and replacement, enable sharing of analysis results with others, and support compliance reporting for critical infrastructure.
* **Costs:** Costs should be minimized by reducing implementation time and staff needs, with efficiency and automation helping justify cybersecurity investments.

**B ) Solution**

Implementing a SIEM system addresses security challenges by centrally collecting, enriching, normalizing, and storing all relevant log data, automatically detecting and reacting to incidents, providing reporting interfaces, and integrating expert knowledge into the analysis process.

**Structure:**

* **Log Collection:** Gathers raw logs from servers, firewalls, applications, and endpoints, providing them to downstream modules.
* **Enrichment:** Adds contextual information to raw logs to improve their quality for better analysis.
* **Normalization:** Converts enriched logs into a standardized format for consistent processing.
* **Correlation and Analysis:** Detects incidents using normalized data, performs real-time and historical analyses, and stores both results and raw logs for future use.
* **Storage:** Acts as a central repository for raw and processed data, supporting historical analysis and compliance reporting.
* **Prepared Data Provision:** Makes processed data available to other modules or systems.
* **Reporting & Threat Intelligence Sharing:** Communicates findings to stakeholders and shares threat intelligence with other systems or organizations.
* **Monitoring & Alerting / Incident Response:** Provides a monitoring interface for analysts and triggers alerts or automated responses for incident management.

**C ) Implementation**

* **Log Collection:** Logs can be pulled from sources via FTP, SCP, or pushed using protocols like syslog or agents, and data streams can be integrated.
* **Enrichment:** Adds contextual information such as IP geolocation, asset discovery, or vulnerability assessment results to make logs easier for analysts to interpret.
* **Normalization:** Converts heterogeneous log formats (relational DBs, plain text, syslog) into a uniform structure to facilitate detection rules and event correlation.
* **Correlation & Analysis:** Conducted in real-time for immediate detection and historically for forensics, using rules that are manual, automatic, or shared, while providing prepared data for downstream modules.
* **Reporting & Threat Intelligence Sharing:** Generates internal and regulatory reports and allows sharing threat intelligence with other organizations, e.g., via AlienVault Open Threat Exchange.
* **Monitoring & Human Interaction:** Analysts verify incidents, distinguish false positives, and create or extend detection rules through visual interfaces for expert decision-making.

**Example 1: IBM QRadar**

* **Data Collection:** QRadar Event Collectors and QFlow Collectors gather and normalize logs.
* **Data Processing:** Enrichment uses network topology (QRadar Risk Manager) and vulnerability data (QRadar Vulnerability Manager), and the Custom Rules Engine analyzes logs; data is stored for forensic investigations.
* **Data Searches / Monitoring:** QRadar Console provides human interaction, report generation, and alerting, implementing all major SIEM modules under brand-specific terminology.

**Example 2: DINGfest Project**

* **Data Acquisition:** Logs collected via Logstash, virtualized systems monitored with VM introspection, normalized into JSON, and streamed via Apache Kafka.
* **Data Analysis:** Enrichment uses Identity Behavior Analytics; events processed using a fingerprint database; Visual Security Analytics supports monitoring and alerting.
* **Digital Forensics & Incident Reporting:** Stores raw data and analysis results in an IoC Vault, allows incident sharing in STIX format, and provides advanced features like VM snapshots, visual analytics, and pseudonymization for privacy.

**Comparison:** DINGfest lacks custom internal reports and incident response features but excels in advanced forensic evidence collection, privacy protection, and enhanced expert interaction.

**D ) Variants :** Not all SIEM systems strictly follow the standard pattern. In some systems, logs are normalized before enrichment, though the input to correlation and analysis remains consistent. Many vendors allow certain modules, such as storage, to be moved to the cloud without significantly altering the overall architecture. Additionally, some SIEMs include active defense features that can mitigate attacks preemptively by leveraging centralized threat intelligence provided by the system.

**E ) Known uses:**

* **IBM QRadar:** Modular system for medium to large organizations; deployable as standalone, distributed, or cloud-based.
* **Splunk Enterprise:** Provides core SIEM functionality (data collection and storage) and is extensible with packages like User Behavior Analytics (UBA).
* **LogRhythm:** Combines host and network monitoring with core SIEM features; includes an AI engine to automate certain operations.
* **McAfee ESM:** Connects with threat intelligence feeds for added context and signatures; supports big data technologies such as Elasticsearch and Kafka.
* **AlienVault OSSIM & USM:** OSSIM is a widely used open-source SIEM, while USM is a paid cloud-hosted version with additional features.

**F ) Consequences:**

* **Detection rate and false positives:** Correlating data from multiple sources and using context information improves detection of complex attacks and reduces false positives.
* **Time of discovery and reaction:** Centralized log collection accelerates detection and enables automatic incident response.
* **Usability:** Centralized interface allows monitoring multiple systems and easy integration of new log sources.
* **Visual data preparation & expert integration:** Monitoring modules provide visually enriched data, allowing experts to refine automated analyses and interpret complex logs.
* **Degree of automation:** Logs are automatically normalized, correlated, and analyzed; reports, dashboards, and incident responses reduce analyst effort.
* **Number of connectible devices:** Interfaces allow connection of any log-generating device, with normalization ensuring consistent processing.
* **Analytics:** Central correlation and analysis detect threats automatically, while human analysts can manually review and analyze data.
* **Reusability & intelligence sharing:** Modular design and standardized interfaces support reusability, threat intelligence sharing, and compliance reporting.
* **Costs:** Automation and centralized monitoring reduce staff requirements and operational costs.

**Liabilities / Challenges:**

* **High effort for introduction:** Integrating SIEM into complex, historically grown IT infrastructures is difficult, and heterogeneous organizational demands complicate standard solutions.
* **Demand for experts:** Experts are still needed for setup, monitoring, and tuning, despite reduced daily workload.
* **Limited human observation input:** SIEM relies on machine-generated logs and context data, so human observations like social engineering reports cannot be directly integrated.

**G ) Related Patterns:**

* **Security Patterns for IDS:** IDS patterns focus on analyzing client requests at specific network points and are not centralized; within SIEM, an IDS can act as a log source for correlation.
* **Centralized Systems Logging Pattern:** Uses the Factory pattern to create loggers for different applications and collects all logs centrally, complementing SIEM but lacking analysis or security-specific processing.
* **Adapter Pattern:** Enables integration of modules with incompatible interfaces post-installation, simplifying connection of external data sources or consumers to the SIEM system.

**Pattern Mining / Identification Process**

**Domain Definition:** Establish the domain of SIEM systems and explain domain-specific terminology.

**Coverage Consideration:** Focus the pattern on one abstraction level, emphasizing software components of SIEM rather than SOC staff processes.

**Information Format Design:** Represent SIEM patterns using UML component diagrams due to their suitability for the abstraction level and familiarity among software engineers.

**Information Collection:** Gather information from multiple sources including software products, research papers, whitepapers, manuals, developer documentation, and product websites to ensure a holistic understanding.

**Information Review:** Analyze advanced and representative SIEM systems using the Gartner Magic Quadrant, considering practical usefulness and strategic vision; systems reviewed include IBM QRadar, Splunk Enterprise Security, LogRhythm Enterprise, HPE ArcSight, McAfee Enterprise Security Manager, and the open-source OSSIM.

**Conclusion and Future Work**

The paper derived a generic SIEM pattern by analyzing advanced SIEM systems and applying a literature-based pattern identification process. Key steps included identifying forces affecting SIEM systems, visualizing the pattern using a UML component diagram, discussing advantages and liabilities, and comparing it with both commercial (IBM QRadar) and research-based (DINGfest) SIEM systems. The pattern helps developers define and modularize SIEM components, improving structure and maintainability, and also serves as a foundation for understanding SIEM functional principles and assembly.

As a high-level framework, the pattern does not cover every detail. Further research is needed to create additional patterns for specific SIEM modules and to break down each module into sub-modules with detailed functionality, ensuring more granular guidance for implementation and advanced system design.

**Security Information and Event Management (SIEM): Analysis, Trends, and Usage in Critical Infrastructures**

**Introduction**

Cybersecurity risks in industrial control systems (ICS/ICT) have increased due to sophisticated attacks by nation-states and cybercriminals, including ransomware, malware, phishing, business email compromise, and data theft through social engineering. To address these threats, SIEM systems play a central role in modern Security Operations Centers (SOC) by collecting and correlating events from sensors such as firewalls, IDS, and antivirus, detecting anomalies, and supporting incident response; for example, NIST recommends ICS SIEMs provide real-time behavioral anomaly detection and intelligent network visualization. The market includes classic providers like HP, IBM, and McAfee, visionary SIEMs such as AT&T Cybersecurity/AlienVault, and promising tools like Splunk. The goal of the research paper is to review SIEM tools, analyze their features and limitations, propose improvements, and study how external factors may impact their future use, particularly in critical infrastructure environments.

**SIEM Solutions**

SIEM systems were created to help administrators design security policies and manage events from multiple sources. A typical SIEM includes source devices that generate logs (firewalls, IDS, servers), log collection, parsing and normalization to standardize formats, a rule engine for applying security rules, log storage, and event monitoring to track suspicious activity; failure of any component, such as log storage, can disrupt the entire system. SIEM capabilities include real-time analysis of security events from networks and applications, and some next-generation SIEMs can automate countermeasure deployment, though often without assessing the full impact of attacks or responses. In the market, SIEMs are classified in Gartner’s Magic Quadrant as leaders, challengers, niche players, or visionaries, with solutions like Splunk and IBM QRadar typically ranked as leaders, while smaller tools fall into niche categories.

**Siem Classification:** SIEM systems are classified based on commercial success and technical capabilities, with annual Gartner reports ranking vendors as leaders (e.g., RSA NetWitness, IBM QRadar, ArcSight, McAfee ESM, LogRhythm), visionaries with strong future plans, challengers performing well today but lacking future vision, and niche players focused on specialized segments, with top vendors including Splunk and AlienVault/AT&T Cybersecurity. Over the past decade, some vendors merged to adapt to market changes, such as IBM with Q1 Labs and HP with ArcSight. Core SIEM capabilities include threat intelligence detection, compliance, and log management, while new innovations encompass UEBA (User and Entity Behavior Analytics), smart dashboards, pre-built reports, incident response workflows, advanced analytics, and correlation searches. Areas for improvement remain in behavioral analysis, risk assessment, visualization, data storage, and automated reaction capabilities, with modern SIEMs increasingly helping administrators detect threats automatically, generate dashboards, and enhance incident response.

**SIEM Features and Capabilities**

* **Correlation rules** – Detect threats by linking events; advanced SIEMs like ArcSight support complex searches.
* **Data sources** – Collect events from diverse sources including sensors and threat intelligence, with QRadar and USM requiring extra components for full coverage.
* **Real-time processing** – Handle millions of events instantly for detection or prevention; all major SIEMs perform well.
* **Data volume** – Manage and analyze large datasets, though storing live data long-term is costly.
* **Visualization** – Custom dashboards and interactive exploration aid understanding of events.
* **Data analytics** – Includes anomaly detection and machine learning; UEBA analyzes user and application behavior.
* **Performance** – Speed of correlation, indexing, searching, and monitoring.
* **Forensics** – Capture and analyze network traffic; ArcSight/LogRhythm capture full sessions, QRadar/Splunk capture selected packets.
* **Complexity** – Ease of deployment and management varies; ArcSight is complex, LogRhythm and Splunk easier.
* **Scalability** – Ability to expand infrastructure to handle more sensors and events.
* **Risk analysis** – Evaluate infrastructure risks natively or via integrations.
* **Storage** – Duration of event retention for future analysis, usually under 90 days.
* **Price** – Licensing and costs vary; enterprise-grade solutions are expensive, while LogRhythm, USM, and SolarWinds are more affordable.
* **Resilience** – Fault tolerance, disaster recovery, and high availability.
* **Reaction and reporting** – Automated incident response, reporting, and alerting.
* **UEBA** – Native or integrated support for user and entity behavior analytics.
* **Security** – Encryption and security automation throughout monitoring, detection, correlation, analysis, and reporting.

**Limitations of Current SIEMs**

* **Incomplete Data:** Most current SIEMs do not capture all relevant sources such as DNS traffic, endpoint logs, or email activity, which limits the system’s ability to correlate all security events. User identities are often fragmented, with shared accounts or multiple roles per user, increasing false positives and negatives. Privacy regulations like GDPR further restrict data usage, requiring next-generation SIEMs to balance comprehensive monitoring with legal compliance. Additionally, current SIEMs rarely provide high-level security risk metrics, making operational decision-making harder.
* **Basic Correlation Rules:** Existing SIEMs rely mainly on simple boolean logic and ad hoc event chaining, which focuses on syntax rather than semantic understanding of attacks. This approach is insufficient for complex, multi-stage, or zero-day attacks. Advanced correlation engines that can analyze historical data, deviations from normal behavior, and multi-source anomalies are generally lacking, limiting the system’s threat detection capabilities.
* **Basic Storage Capabilities:** Archived data is typically handled manually, with diverse storage solutions ranging from attached disks to in-house distributed file systems (e.g., HDFS) or cloud platforms. Storage duration is often limited (e.g., 6 months), which can hinder the investigation of advanced persistent threats. Manual management of archival data increases operational costs and introduces potential security and reliability risks. Future SIEMs need secure, elastic, and customizable storage solutions that support long-term retention and compliance requirements.
* **Reliance on Humans:** Despite automation in detection and partial response, human analysts are still essential to interpret threats, define countermeasures, and deploy responses. This reliance slows down incident handling, increases costs, and remains error-prone. Effective automation must ensure secure, timely, and impactful responses without replacing human expertise entirely, supporting decision-making rather than fully substituting it.
* **Limited Reaction and Reporting:** Most SIEMs provide only basic automated actions, such as sending emails, generating tickets, or executing scripts, and lack pre-configured workflows for regulatory compliance or incident mitigation. Integration with layered defense strategies (e.g., multiple IDSs, diverse monitoring systems) is minimal, limiting accuracy and increasing false alarms. Reporting is often generic, not following standards required by supervisory authorities, reducing its usefulness for governance and audit purposes.
* **Limited Data Visualization:** Visualization features in current SIEMs are often basic, generic, or too rudimentary to support complex analysis. They rarely integrate diverse data types like OSINT, statistical modeling, or user behavior models. Analysts have difficulty extracting actionable insights or profiling attacks effectively. Future SIEMs must support flexible, real-time visualizations that handle streaming data, correlate heterogeneous sources, and communicate attack provenance, ongoing activities, vulnerabilities, and user/session behavior.

**The Future of SIEMs**

* **Political Factors:** Political initiatives increasingly influence the development and adoption of SIEM systems. Governments are emphasizing the protection of sensitive business and personal information, creating security frameworks and regulations that can drive investment in cybersecurity technologies. For example, the European Union has committed substantial funds to strengthen cyber resilience, including public-private partnerships that foster innovation in security solutions. These political measures can empower SIEM adoption by ensuring regulatory support and promoting standardized, high-quality, and interoperable cybersecurity products.
* **Economic Factors:** Economic trends such as the rise of temporary work, freelance employment, and globalization directly impact the SIEM market. Companies increasingly require tools that are user-friendly and adaptable to a diverse workforce, including freelancers and short-term employees who may use unmanaged devices. The rapid growth in cybersecurity jobs and the expanding threat landscape for small and medium-sized enterprises create opportunities for SIEM providers, particularly for cloud-based or “SIEM-as-a-Service” models. Large multinational organizations also drive demand for scalable solutions capable of monitoring global networks efficiently.
* Societal Factors: Societal reliance on information and communication technology (ICT) is growing, especially among younger, tech-savvy generations. These users understand the importance of cybersecurity, influencing organizational awareness and behavior. The rapid expansion of social media platforms introduces additional data sources for security monitoring, while the increasing sophistication of cyber-attacks—targeting both organizations and critical infrastructure—underscores the essential role of SIEM systems. However, the deep web presents challenges in data collection and event analysis, requiring SIEMs to adapt to more complex information environments.
* **Technological Factors:** Emerging technologies significantly shape the evolution of SIEMs. Cloud storage and cloud service integration enable scalable, high-availability analytics platforms, while mobile devices and BYOD trends introduce new security risks. Advanced technologies such as big data analytics, machine learning, the Internet of Everything (IoE), 5G networks, and social media analytics enhance SIEM capabilities for real-time detection, anomaly identification, and intelligent decision-making. However, these advancements also increase data volumes and complexity, requiring SIEMs to adapt in terms of storage, processing, and automated response mechanisms.
* **Legal Factors:** Data protection regulations, such as the EU’s General Data Protection Regulation (GDPR), directly affect how SIEMs collect, process, and store data. Regulations treat IP addresses and other user-related information as personal data, requiring compliance in terms of secure storage, access control, and retention policies. SIEM systems must balance effective security monitoring with legal constraints on user privacy, ensuring that sensitive data is handled responsibly while maintaining operational effectiveness in detecting and responding to incidents.
* **Environmental Factors:** The proliferation of cloud services, mobile devices, and IoT endpoints is creating new challenges for SIEM systems. While traditional computers are harder to compromise, connected devices like cameras, appliances, and industrial sensors remain vulnerable, increasing the number of potential attack vectors. Small and medium businesses are increasingly adopting cloud-based SIEM modules, which help manage data storage and security operations cost-effectively. Future SIEMs must evolve to efficiently collect, organize, and analyze the rapidly growing volume of events generated by these diverse endpoints and technologies.

**Potential Enhancements of Future SIEMs**

* **Diverse Security:** Future SIEMs should enhance system security by integrating diversity-related technologies. This involves evaluating the similarities and differences between security mechanisms, vulnerabilities, and attack patterns, which is currently underexplored in research. Quantitative and probabilistic metrics can guide decisions on combining multiple defenses effectively. By assessing the collective strengths and weaknesses of diverse security layers, SIEMs can achieve improved resilience. Rather than only aggregating diverse machine learning models, next-generation SIEMs should focus on integrating diverse inputs from multiple monitoring sources to enhance detection and response accuracy.
* **OSINT Data Fusion:** Integrating Open-Source Intelligence (OSINT) is a potential enhancement for SIEMs. By applying natural language processing, SIEMs can analyze keywords across multiple languages to identify threats, such as DDoS attacks, leaks, or breaches. This process not only tags relevant OSINT data but can also extract additional information like locations, entities involved, and threat context. Including prediction confidence scores helps reduce false positives and provides security analysts with more reliable, actionable intelligence for decision-making.
* **Enhanced Visualization:** Visualization capabilities are crucial for understanding complex security data. Future SIEMs should develop advanced visualization models capable of handling high-dimensional, temporal, textual, relational, and spatial data. Features such as interactive exploration, comparative analysis, and real-time streaming visualizations will enable better situational awareness. Specialized visual summaries of user behavior and system activity can reveal common patterns, anomalies, and potential threats, helping analysts focus on critical security events efficiently.
* **Enhanced Storage:** Reliable and secure data storage is essential for long-term SIEM effectiveness. Future systems should provide flexible and elastic archiving solutions, leveraging cloud services like Amazon S3, Glacier, or Azure Blob Store. These solutions must allow customizable retention policies, secure storage, and efficient handling of large-scale historical data, ensuring accessibility for forensic analysis and compliance purposes while minimizing risks and costs associated with traditional on-premises storage.
* **Integration with SOAR:** Security Orchestration, Automation, and Response (SOAR) solutions complement SIEMs by automating workflows, repetitive security tasks, and incident management. Next-generation SIEMs should integrate adaptive SOAR capabilities to dynamically interact across all phases of incident response. When combined with Threat Intelligence Platforms (TIPs), Endpoint Detection and Response (EDR), and Next-Generation Firewalls (NGFW), SIEM-SOAR integration enables proactive detection, prevention, and response to complex cyber threats. Examples include adaptive frameworks from Splunk, Nokia NextGuard, and Integrated Adaptive Cyber Defense (IACD).
* **AI/ML Capabilities:** Artificial intelligence and machine learning are key enablers for smarter SIEM operations. AI/ML integration enhances threat detection, correlation, and automated reaction capabilities by identifying abnormal patterns in network traffic, applications, and user behavior. Next-generation SIEMs should combine rule-based analysis with AI-driven behavioral models, allowing anomaly detection, prioritization of critical assets, and automated responses. Unsupervised statistical learning can establish baselines for normal behavior and highlight deviations, improving detection accuracy and reducing false positives.
* **Other Potential Enhancements:** Future SIEMs should develop advanced risk-based metrics across multiple operational layers, including hosts, applications, and services, to provide high-level security insights. The rise of 5G and IoT technologies necessitates hierarchical and collaborative SIEM architectures capable of processing increased data volumes closer to the network edge. Integration with Extended Detection and Response (XDR) platforms can provide complementary benefits, focusing on compliance, threat detection, and operational risk. Overall, next-generation SIEMs must emphasize autonomy, simplified deployment, and reduced operational effort to lower costs while maintaining comprehensive security coverage.

**SIEMs in Critical Infrastructures**

* **Overview of Critical Infrastructures and Security Challenges:** Critical infrastructures (CIs) are organizational and physical systems whose failure can disrupt public safety, national security, and essential services. They rely on Supervisory Control And Data Acquisition (SCADA) systems and Industrial Control Systems (ICS) to manage complex processes across sectors such as energy, water, transport, healthcare, and finance. Unlike conventional IT networks, CI networks operate in real time at the physical level, emphasizing system availability over confidentiality. Many ICS are legacy systems with limited security features, making them vulnerable to cyber-attacks. Traditional vulnerability assessments are often infeasible as they could disrupt operations. SIEM systems are essential for monitoring, correlating, and analyzing massive amounts of data to detect threats, support mitigation strategies, and provide real-time situational awareness across CI networks.
* **Energy Distribution:** The energy sector, including power grids, gas, and oil, is particularly susceptible to cyber-attacks due to its large attack surface, the increasing number of threat actors, and complex interdependencies between physical and cyber systems. Threats range from billing fraud through smart meters to potential physical destruction. SIEM solutions help monitor network signals, detect anomalies such as GPS spoofing or DDoS attacks, and alert security analysts to mitigate risks. By providing detailed, real-time correlation of events, SIEMs enhance the protection of critical assets and operational continuity in the energy sector.
* **Water Supply:** Water infrastructure faces threats at every stage, from source to distribution. SIEM systems monitor SCADA networks in real time, detecting deviations in water quality or unusual activities that may indicate an attack. Examples include network reconnaissance, malware, man-in-the-middle attacks, or changes in operator behavior. Defense-in-depth strategies, supported by SIEMs, are critical to understand interconnections between vulnerabilities, threats, and mitigations. Advanced SIEMs enable IT and OT systems to work together, automate incident responses, and bridge OT security gaps, ensuring compliance and efficient infrastructure protection.
* **Transportation and Healthcare:** The digitization of transportation networks and healthcare infrastructures has increased cyber risks. Transportation systems face threats ranging from ransomware and data breaches to disruption of physical assets, affecting aviation, highways, maritime, and railways. SIEM solutions support cybersecurity planning, threat detection, and employee awareness while monitoring both digital and physical assets. In healthcare, the adoption of cloud-based electronic health records, personal medical devices, and data sharing between stakeholders increases vulnerability to malware, ransomware, and phishing attacks. SIEMs provide real-time analytics, scalable log management, compliance reporting, and a consolidated view of all security events, protecting sensitive patient data and ensuring regulatory compliance.
* **Financial Services:** Financial institutions are prime targets for both external and insider threats, including data exfiltration, fraud, and account abuse. Challenges include business scaling, compliance with diverse legal frameworks, and detection of insider threats. Modern SIEMs enhance security through User and Entity Behavioral Analytics (UEBA), which detect deviations from baseline behaviors, monitor abnormal network traffic, and support automatic regulatory reporting. By streamlining incident reporting, detecting anomalies, and providing forensic capabilities, SIEM solutions help financial organizations protect assets, ensure compliance, and respond proactively to both cyber and operational risks.

**Related Work**

SIEM systems can be evaluated from commercial and technical perspectives, each offering distinct insights. Commercial evaluations by Gartner, Techtarget, Info-Tech, and Solutions Review focus on vendor landscape, market trends, adoption, and ROI. Gartner’s Magic Quadrant categorizes vendors, while Info-Tech and Solutions Review compare benefits, drawbacks, and compliance features. However, these analyses rarely consider technical enhancements or external factors affecting long-term SIEM evolution, limiting guidance for designing next-generation solutions.

Technical evaluations emphasize internal capabilities and improvements. Caccia et al. suggest User and Entity Behavior Analytics (UEBA) for better threat detection, while Kotenko and Chechulin propose attack modeling, internal security repositories, and advanced security metrics for IoT-based SIEMs. Yet, they often ignore commercial and external factors. A holistic approach integrating commercial features, technical capabilities, and PESTLE influences can guide the development of next-generation SIEMs with improved analytics, SOAR integration, and adaptability to evolving threats.

**Conclusion**

The paper evaluates eight leading SIEM solutions—ArcSight, QRadar, McAfee, LogRhythm, USM OSSIM, RSA, Splunk, and SolarWinds—based on market performance and technology evolution. Key observations highlight the need for advanced behavioral and risk analysis with multi-level metrics and diverse sensor deployment. Current GUIs are user-friendly, but visualization struggles with large event volumes, and reaction capabilities remain limited, requiring enhancements for faster decision-making and threat response.

Data storage is generally sufficient, but local hardware limits and licensing costs for large volumes are challenges. Future SIEMs could use secure, elastic cloud storage with customizable retention policies. They will remain crucial for industrial control systems, SOCs, and SMEs, while considering PESTLE factors—political, economic, social, technological, environmental, and legal—to guide investment, enhance capabilities, and address evolving cybersecurity challenges.

**Challenges and Directions in Security Information and Event Management (SIEM)**

**Introduction**

Situational awareness (SA) is the ability to collect, process, and fuse data to understand their technical meaning and make informed decisions. In cybersecurity, SA relies on multiple sources such as intrusion detection systems, network audits, integrity monitors, and system or application logs, often processed hierarchically to extract meaningful insights. Security Information and Event Management (SIEM) systems are central to this process, collecting, normalizing, and analyzing diverse security data, and serving as the core tool in Security Operations Centers (SOCs). However, SIEM deployment is challenging, requiring ad-hoc data collectors, correlation rules, and significant human expertise to handle large data volumes and unknown attacks.

Industrial experiences, such as in Air Traffic Control systems by LEONARDO, show that integrating SIEM with tools like OSSEC and building behavioral baselines can enhance security. By profiling normal system behavior and detecting anomalies outside these baselines, organizations can identify threats beyond predefined correlation rules. Advanced techniques like Latent Dirichlet Allocation (LDA) enable pattern recognition in unstructured logs, supporting anomaly detection and data-driven discovery. Such approaches complement traditional SIEM methods, improving situational awareness and the overall effectiveness of cybersecurity operations.

**Reference System**

The NAPOLI FUTURA project (“Novel Approaches to Protect Critical Infrastructures from Cyber Attacks”) is an industry-academia collaboration involving LEONARDO and Critiware S.r.l., focusing on integrated monitoring, virtual machine live migration, and security data analytics using Apache Storm. The project provided hands-on experience with real operational data, such as flight simulations, where researchers monitored logs and detected unusual access patterns in real time through analytics.

The Air Traffic Control (ATC) system architecture consists of distributed nodes (D02, MN1, SFN) supporting critical functions like flight progress monitoring (MONT), conflict detection (MTCD), and controller working positions (CWP). Logs are massive (~15,000 lines per minute) and heterogeneous, coming from ~20 sources across 7 nodes, often unstructured text, including standard syslogs and legacy application logs. The main challenge for SIEM is to correlate, normalize, and analyze these diverse logs to detect anomalies and potential security incidents, filtering noise while highlighting unusual patterns effectively.

**Challenges in SIEM**

Modern SIEMs like LogRhythm and Splunk ES face challenges with data formats, as they handle structured logs efficiently through internal representations (e.g., MDI Fabric or CIM), but unstructured text logs are imported mostly as raw messages. While fields such as timestamps or hostnames are recognized automatically, free-text messages like “WARNING I received zero on 'address'. Converting to default 'DUMMYADDRESS'” remain unparsed, making automated analysis difficult.

Correlation rules also have limitations, requiring analysts to define them manually (e.g., “alert if failed login > 5”), which leaves unknown multi-source attacks potentially undetected. To address this, behavioral baselines can be built using NLP and information retrieval techniques to model normal system behavior. By comparing runtime log messages against these baselines, actions outside regular patterns—such as unusual flight updates or system warnings—can be flagged automatically, complementing traditional rule-based detection.

**Explorative Integration of ATC-SIEM**

OSSEC is an open-source tool combining host-based intrusion detection (HIDS), log monitoring, and SIEM functionality, and is featured in OSSIM (AlienVault’s open-source SIEM). It provides built-in alerts for standard services like Apache, FTP, and syslog, with detection rules stored in XML files. For example, OSSEC can automatically alert if Apache logs show multiple failed login attempts.

Adapting OSSEC for ATC systems required handling proprietary, unstructured logs. A connector program was developed to poll logs, wrap each line into standard syslog format with fields like timestamp, hostname, and application, and forward them to the OSSEC server. Domain-specific rules were added in local\_rules.xml, such as raising an alert when the keyword “calcnextsecgeog” appears. This approach leverages OSSEC’s monitoring infrastructure while enabling custom SIEM correlation rules for proprietary ATC logs.

**Modeling Behavioural Baseline**

**Model Construction Using LDA:** Traditional SIEMs rely on hard-coded correlation rules, which cannot anticipate unknown attacks, especially in ATC systems with unstructured logs. To address this, Latent Dirichlet Allocation (LDA) was used to model normal system behavior. Logs were split into 10-second chunks, parsed to remove variables, and converted into a chunk-term matrix. LDA was trained using 10-fold cross-validation, and the optimal number of topics (e.g., K=12 for D02-PAN) was selected based on the “knee” of the perplexity curve. Topics identified recurring operations, such as flight updates or timers, and visualization via LDAvis highlighted frequent system behaviors.

**Compromise Detection:** Deviations from the LDA-based behavioral baseline were used to detect anomalies. Each log chunk’s perplexity was computed against the normative model, with high perplexity indicating potential unknown or anomalous actions. Experiments with synthetic anomalies at 1%, 5%, and 10% of log lines showed that even low-level anomalies achieved high recall (0.93) and precision (0.90), while higher percentages were detected perfectly. This unsupervised, fully automated approach scales to high-volume unstructured logs and complements traditional SIEM rules by detecting previously unknown anomalies, such as unusual flight data updates.

**Conclusion and Future Work**

The practical implementation of SIEM in a real-world Air Traffic Control (ATC) system demonstrated the challenges of managing high-volume, unstructured logs and the absence of predefined threat models. Key contributions included the development of behavioral baselines using LDA for anomaly detection, the integration of OSSEC to monitor proprietary logs, and the combination of data-driven approaches with traditional SIEM correlation rules. This approach enhanced the detection of unknown anomalies while maintaining the benefits of existing SIEM infrastructures.

Future enhancements aim to expand SIEM capabilities through plugins, enabling improved monitoring and correlation across diverse log sources. Advanced correlation techniques will be explored to detect stealthy, multi-step attacks spanning multiple nodes or logs. Visualization tools will be developed to help analysts interpret large volumes of data efficiently. Additionally, automated cognitive analysis is proposed to mimic human reasoning, aiming to improve detection efficiency and reduce the dependency on human expertise.

🤖 Prototypes

Dataset link:

[GitHub - logpai/loghub: A large collection of system log datasets for AI-driven log analytics [ISSRE'23]](https://github.com/logpai/loghub)

[Windows\_2k.log\_structured.csv](https://drive.google.com/open?id=1M7OClPgx_5VoYr71kZ35tuuCz5Ynj7wg)

[Mac\_2k.log\_structured.csv](https://drive.google.com/open?id=18-v5pbBnsbRGsSw939AhicWY3GPMwGw3)

**Prototype:**

[Prototype-001](https://colab.research.google.com/drive/1vIqEI533J1e_qqTBpCKjoKTl1shRLeT-#scrollTo=gjrl8KEbVBR8)