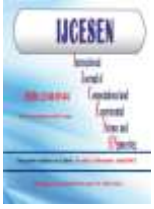
International Journal of Computational and Experimental Science and ENgineering
(IJCESEN)

Vol. 11-No.4 (2025) pp. 9295-9301
http://www.ijcesen.com

ISSN: 2149-9144



Research Article

Federated Security Control Data Fabric: Scalable Telemetry Normalization and Orchestration in Multi-Cloud Environments

Karthikeyan Thandayutham*

Independent Researcher, USA
* **Corresponding Author Email:** karthikeyant.thandayutham@gmail.com - **ORCID:** 0000-0002-5247-9990

Article Info:

DOI: 10.22399/ijcesen.4414
Received : 29 September 2025
Revised : 20 November 2025
Accepted : 29 November 2025

Keywords

Multi-Cloud Security Architecture,
Federated Data Fabric,
Security Telemetry Normalization,
AI-Driven Threat Prioritization,
Distributed Security Operations

Abstract:

The adoption of multi-cloud architectures has fundamentally reshaped enterprise security operations, introducing unprecedented complexity in managing controls across heterogeneous environments. Traditional security paradigms built around centralized log collection through Security Information and Event Management (SIEM) and Security Orchestration and Automation Platforms(SOAR) struggle to scale economically and operationally when confronted with distributed cloud-native telemetry. Organizations now operate across multiple public cloud providers, each emitting high-volume preventive, detective, and remediative control telemetry in proprietary schemas. The financial cost of data egress and the processing delays introduced by centralized aggregation undermine real-time threat detection, while provider-specific visibility creates blind spots that sophisticated adversaries can exploit.

This paper proposes the Federated Security Control Data Fabric (F-SCDF) as a distributed-first architectural framework for multi-cloud security telemetry. The fabric keeps telemetry processing close to its source while enabling unified semantic interpretation and cross-environment orchestration. Distributed ingestion gateways perform local normalization, enrichment, and filtering to minimize data egress and accelerate event availability. A universal security control schema and semantic mapping registry provide consistent interpretation of heterogeneous events, while a federated lakehouse architecture enables unified querying across source environments without requiring centralized storage. An AI-driven signal prioritization broker applies machine learning models to suppress noise, risk-score events, and feed downstream orchestration systems with actionable intelligence rather than raw alert volume.

The F-SCDF architecture delivers vendor-agnostic security operations that align economic efficiency with real-time detection requirements, providing a scalable foundation for protecting distributed enterprise assets as multi-cloud complexity and telemetry volume continue to grow.

1. Introduction

The modern enterprise technology environment has been fundamentally transformed by the introduction of multi-cloud architectures, which have radically altered how organizations manage security operations and controls.The Flexera 2024 State of the Cloud Report states that multi-cloud strategies have become a central part of organizational digital infrastructure and that organizations are operating in complex environments that cut across a variety of cloud providers as well as on-premises cloud deployments[1]. The strategic change is an indication of the perception that no single cloud provider can best meet all the business needs, and thus organizations end up distributing workloads based on the different platforms, depending on the capabilities, costs, and regulatory needs.

The security considerations of this distributed architecture are far-reaching and many-faceted. Conventional security frameworks, which are based on centralized aggregation and analysis of logs, are faced with more challenges than ever before due to the heterogeneity of multi-cloud telemetry. As Gartner has emphasized in their cloud spending forecasts, the rapid expansion of public cloud adoption drives substantial increases in infrastructure complexity, with security and risk management becoming critical investment priorities for organizations seeking to protect distributed assets [2]. The fundamental challenge lies in reconciling comprehensive visibility requirements with the operational realities of managing security controls across environments that employ different logging formats, event taxonomies, and security primitives.

This paper introduces the Federated Security Control Data Fabric as an architectural response to these challenges, proposing a distributed approach that maintains telemetry proximity to its source while enabling unified visibility and orchestration. Unlike centralized security information and event management systems that aggregate all logs into a single repository, the proposed fabric employs distributed ingestion gateways, semantic normalization, and intelligent signal prioritization to achieve scalable multi-cloud security operations. By leveraging federated data management principles combined with artificial intelligence-driven event prioritization, this architecture addresses both the economic constraints imposed by data egress costs and the operational imperatives of real-time threat detection across heterogeneous cloud environments.This work is primarily architectural and conceptual; it synthesizes existing technologies into a unified design pattern rather than presenting empirical benchmarks.

2. The Multi-Cloud Security Telemetry Challenge

Beyond the economic challenges, the technical heterogeneity of multi-cloud environments creates profound obstacles to effective security correlation and analysis. Each cloud provider employs proprietary event schemas, logging formats, and semantic definitions for security-relevant activities. An authentication failure in one provider's identity and access management system may be represented with entirely different field names, severity indicators, and contextual metadata compared to an equivalent event in another provider's environment. Comparative studies of performance and security in distributed systems have highlighted how these architectural variations complicate efforts to establish unified security monitoring, as security teams must maintain expertise in multiple provider-specific toolsets while manually correlating events that may represent coordinated attack activity spanning multiple cloud boundaries [5]. The absence of standardized representation of events between providers implies that the security analysts will find the patterns and anomalies difficult to see across different heterogeneous environments, and this gives room for the advanced adversary to take advantage of any visibility gaps.The latency characteristics of centralized log aggregation contributes to the inefficiency of real-time security operations. When security telemetry must traverse multiple network boundaries, potentially crossing geographic regions and provider networks, the resulting delays can prove critical in time-sensitive scenarios such as detecting and blocking credential compromise or preventing lateral movement following initial access. Research examining security implementations in cloud environments has demonstrated that processing delays introduced by centralized architectures can significantly impact the mean time to detect and respond to security incidents, particularly for attacks that employ rapid automated techniques [6]. These latency considerations become even more critical as organizations seek to implement automated response mechanisms that depend on near-real-time event analysis to effectively contain threats before they can achieve their objectives.

3. Limitations of Existing Security Architecture Models

Contemporary security architectures employed in enterprise environments typically rely on centralized Security Information and Event Management platforms that aggregate logs from distributed sources into unified data repositories for correlation and analysis. While this model proved effective in traditional on-premises data center environments with manageable data volumes, it encounters severe scalability limitations when applied to multi-cloud contexts. The fundamental assumption underlying centralized SIEM architectures—that aggregating all security-relevant data into a single location provides optimal visibility—breaks down when confronted with the economic realities of cloud data egress pricing and the sheer volume of telemetry generated by modern cloud-native applications. Organizations implementing centralized log aggregation discover that the costs of transferring data out of native cloud environments, combined with the storage and processing expenses of SIEM platforms, can quickly escalate to unsustainable levels [7].

The operational challenges extend beyond mere cost considerations. Centralized SIEM systems demand constant maintenance of custom parsing rules and normalization logic of every individual source of logs emitted into the system. The integration burden increases in proportion to the number of new cloud services picked up by organizations, new security tools deployed, as well as the workloads being moved into new providers. Security teams end up spending a lot of time creating and maintaining such integrations as opposed to the threat detection and investigation efforts. Research examining security operations in distributed computing environments has documented how this integration complexity creates operational friction that slows the adoption of new technologies and services, as security teams must evaluate not only the security implications of new tools but also the effort required to integrate their telemetry into existing monitoring infrastructure [8].

Another way that has emerged is cloud-native security tools that are offered by the major cloud vendors, which offer extensive integration with platform-specific services and utilize provider-specific threat intelligence. Such tools as AWS GuardDuty, Azure Sentinel, and Google Cloud Security Command Center provide advanced threat detection features tailored to the ecosystem of the products. This is because of the single-cloud focus, which introduces serious visibility gaps in the multi-cloud environment. Each provider's security tooling operates independently, employing proprietary event schemas and threat models that do not naturally interoperate with other providers' offerings. Studies examining machine learning applications in cloud security have highlighted how this fragmentation prevents organizations from developing unified behavioral baselines or detecting attack patterns that span multiple cloud environments [9]. The security analysts should acquire proficiency in various provider-specific interfaces and query languages and have to switch context between various security consoles at any given time in the course of an investigation.

Security Orchestration, Automation, and Response platforms have their limitations that make these problems even greater. SOAR systems promise to automate routine security operations through pre-defined playbooks that execute standardized response procedures when specific alert conditions are met. However, these platforms inherit the centralization challenges of SIEM architectures while introducing additional constraints around adaptability. SOAR playbooks operate on rule-based logic that struggles to account for contextual factors such as asset criticality, user behavior patterns, or evolving threat landscapes. Research examining cybersecurity protection mechanisms has demonstrated that rigid automation frameworks can actually introduce new risks when they execute inappropriate responses based on false positive alerts or fail to adapt to novel attack techniques that fall outside predefined playbook scenarios [10]. The reactive nature of SOAR systems means they respond only to alerts that have already been generated and passed through detection thresholds, missing opportunities for proactive threat hunting or early-stage attack interdiction.

4. The Federated Security Control Data Fabric Architecture

The Federated Security Control Data Fabric architecture reinvents the concept of how security telemetry is gathered, standardized, and offered to be analyzed and orchestrated in multi-clouds. The F-SCDF model does not consolidate all of the security logs to a central repository and instead keeps them distributed within the environments where those logs originate, but offers the ability to interpret them semantically and to make queries out of this distributed data as one. This architectural approach directly addresses the economic and operational limitations of centralized models by minimizing data egress, reducing storage costs, and enabling local preprocessing that filters noise before it consumes network and storage resources. The core innovation lies in combining several complementary technologies into an integrated system specifically optimized for security control telemetry management [3].

At the foundation of this architecture are distributed ingestion gateways deployed within each cloud environment and major service boundary. These lightweight components perform critical preprocessing functions at the edge of the telemetry pipeline, transforming provider-specific log formats into standardized events that conform to a universal security control schema. By executing normalization logic locally before data egress occurs, these gateways eliminate the need for downstream systems to maintain provider-specific parsing rules while simultaneously reducing the volume of data that must traverse cloud boundaries. Research on performance optimization in distributed systems has demonstrated that edge preprocessing can substantially reduce both network overhead and central processing requirements, enabling more scalable architectures that maintain effectiveness as telemetry volumes grow [5]. The gateways also perform initial enrichment, appending contextual metadata such as asset classifications, business unit associations, and compliance framework mappings, while this information is readily available in the source environment.

The semantic mapping registry maintains the translation rules that map provider-specific events into a universal security control schema, which defines the standardized vocabulary through which security events can be consistently interpreted regardless of source. Based on principles informed by the new standards, like the Open Cybersecurity Schema Framework, this schema establishes extensive taxonomies of authentication events, authorization decisions, network flows, configuration changes, and threat detections. The mapping registry maintains translation rules that convert provider-specific event formats into this universal representation, creating a layer of abstraction that insulates downstream security operations from the details of individual cloud provider implementations. Studies examining security architectures in cloud computing have emphasized the importance of semantic standardization for enabling effective cross-platform security monitoring, as inconsistent event representation fundamentally undermines efforts to correlate activities and detect distributed attack patterns [6].

The federated lakehouse architecture component enables the F-SCDF to maintain data distribution while providing centralized visibility. Rather than physically aggregating logs, this approach stores raw telemetry in cost-effective object storage within each source environment, indexed with rich metadata that enables discovery and federated querying. Security analysts can search across these distributed repositories using a unified query interface, with the federated query engine pushing computation to data rather than moving massive volumes of raw logs for central processing. This architecture leverages advances in distributed query processing to deliver visibility comparable to centralized systems while preserving the economic advantages of keeping data in its native environment. Research on distributed computing architectures has validated that properly implemented federated query systems can achieve query performance within acceptable ranges of centralized alternatives while dramatically reducing storage costs and data transfer requirements [7].

At the intelligence layer, the F-SCDF incorporates an artificial intelligence-powered signal prioritization broker that applies machine learning models to filter, enrich, and risk-score security events before they reach orchestration and response systems. This component addresses the fundamental signal-to-noise challenge that undermines traditional security operations, where security teams struggle to identify genuine threats amid vast volumes of routine events and false positive alerts. The broker employs unsupervised learning algorithms for anomaly detection, identifying behavioral deviations that may indicate novel attack techniques not captured by signature-based detection systems. It also implements supervised classification models trained on historical incident data to predict which alerts warrant immediate investigation versus those that can be safely deprioritized. Research examining machine learning applications in cloud security has demonstrated that intelligent event prioritization can dramatically reduce analyst workload while improving threat detection effectiveness, enabling security operations centers to maintain effectiveness even as telemetry volumes continue their exponential growth [9]. The broker's risk scoring capabilities integrate multiple contextual factors, including asset criticality, user risk profiles, threat intelligence indicators, and environmental context to produce normalized risk assessments that guide both automated response and human investigation priorities.

5. Implementation Feasibility and Technical Foundation

The practical viability of the Federated Security Control Data Fabric rests upon established technologies and architectural patterns that have already demonstrated scalability and effectiveness in adjacent problem domains. Federated data management approaches have achieved widespread adoption in large-scale analytical systems, where organizations have successfully implemented distributed query processing across heterogeneous data sources while maintaining acceptable performance characteristics. The OpenTelemetry project provides a reference implementation for distributed telemetry collection with local preprocessing, demonstrating that lightweight agents can effectively normalize and enrich observability data before transmission to central systems. These precedents validate the core technical premises of distributed ingestion and semantic normalization that underpin the F-SCDF architecture [8].

Modern lakehouse architectures combine the scalability of data lakes with the query performance and ACID transaction guarantees of data warehouses. Lakehouse architecture has become a mature field with several commercial and open-source systems operating petabyte-scale systems in the production space. Apache Iceberg and Delta Lake are examples of technologies used to offer metadata management and distributed transaction coordination that can be used to provide federated query processing without compromising data consistency. These platforms demonstrate that organizations can achieve centralized visibility without centralized storage, querying across distributed repositories with performance characteristics approaching those of monolithic data warehouses. The economic advantages of this approach prove particularly compelling for security use cases, where retention requirements often mandate preserving telemetry for extended periods while analysis typically focuses on recent time windows. Organizations are able to meet compliance needs by storing historical data in cost-effective object storage with metadata that allows selective querying of the metadata without paying the storage cost that comes with the maintenance of all the historical data in high-performance SIEM systems [4].

Machine learning in security operations has moved beyond research and development, and leading vendors of security platforms have added AI-powered threat detection, alert triage, and automated investigation services. Unsupervised learning anomaly detection algorithms have been successfully used to detect behavioral anomalies that are indicative of insider threats, compromised credentials, and new attack methods. Supervised classification models trained on historical security incident data successfully predict which alerts represent genuine threats warranting immediate investigation. Research examining machine learning deployment in cloud security contexts has documented substantial improvements in both detection accuracy and operational efficiency when intelligent prioritization filters the overwhelming volume of security events generated by modern cloud environments [9]. These implementations validate that AI-powered signal processing can operate effectively at the scale required for enterprise security operations.

The organizational implementation pathway for F-SCDF adoption emphasizes incremental deployment that allows enterprises to realize benefits progressively while managing change risk. Initial phases focus on deploying ingestion gateways in pilot environments to validate normalization schemas and quantify cost reductions from reduced data egress and filtering of low-value events. Subsequent phases introduce federated lakehouse storage for new telemetry streams while maintaining existing SIEM integrations, allowing side-by-side validation of query capabilities and analyst workflows. Organizations then incorporate the AI-powered prioritization broker to demonstrate alert volume reduction and improved investigation efficiency before gradually transitioning orchestration and response systems to consume telemetry through the federated access layer. Studies examining cybersecurity architecture evolution have emphasized that successful transformation initiatives require this type of phased approach that proves value incrementally rather than demanding wholesale replacement of existing infrastructure [10]. This implementation strategy allows security teams to develop expertise with new tools and processes while maintaining operational continuity throughout the transition.

Table 1: *Multi-Cloud Security Telemetry Challenges [3, 4]*

Challenge Category	Description	Impact
Economic Burden	Cloud providers impose substantial data egress charges for transferring logs out of native environments for centralized analysis	Organizations forced to choose between comprehensive visibility and budget constraints
Technical Heterogeneity	Each cloud provider employs proprietary event schemas, logging formats, and semantic definitions	Security teams cannot easily identify patterns across heterogeneous environments
Latency Issues	Security telemetry traversing multiple network boundaries introduces processing delays	Critical time-sensitive scenarios impacted, including credential compromise detection
Volume Growth	Exponential increase in security-relevant log data across multi-cloud environments	Traditional architectures cannot scale to accommodate growing telemetry streams

Table 2: *Limitations of Existing Security Architecture Models [5, 6]*

Architecture Type	Key Limitations	Operational Consequences
Centralized SIEM Platforms	High data egress costs, processing latency, and continuous maintenance of custom parsing rules	Organizations implement selective logging, creating potential security gaps
Cloud-Native Security Tools	Vendor lock-in, incompatible schemas across providers, and fragmented visibility	Analysts must context-switch between multiple consoles during investigations
SOAR Platforms	Reactive alert processing, rule-based rigidity, and centralization dependency	Inappropriate responses from false positives, missed novel attack techniques
Traditional Correlation	Cannot effectively correlate events across cloud boundaries	Attack patterns spanning multiple environments remain invisible

Table 3: *Federated Security Control Data Fabric Architecture Components [7, 8]*

Component	Function	Benefit
Distributed Ingestion Gateways	Transform provider-specific logs into standardized events, perform local preprocessing and enrichment	Minimizes data egress, reduces central processing burden, accelerates event availability
Universal Schema & Mapping Registry	Provides standardized vocabulary for security events across heterogeneous sources	Enables consistent interpretation, simplifies integration of new services
Federated Lakehouse Architecture	Stores raw telemetry in cost-effective object storage within source environments	Dramatically reduces storage costs while maintaining centralized visibility
AI-Powered Signal Prioritization Broker	Applies machine learning for filtering, enrichment, and risk-scoring	Reduces analyst workload, improves threat detection effectiveness
Policy-Aware Federated Access Layer	Enforces least-privilege principles and data residency compliance	Maintains security and regulatory compliance across the distributed architecture

Table 4: *Implementation Feasibility and Technical Foundation [9, 10]*

Technology Foundation	Maturity Level	Application to F-SCDF
Federated Ingestion Patterns	Proven through OpenTelemetry and distributed logging systems	Validates distributed collection with local preprocessing
Lakehouse Architectures	Production deployments at petabyte scale	Enables distributed storage with centralized query capabilities
Semantic Normalization Standards	Emerging through OCSF and industry collaboration	Provides a foundation for a universal security event schema
Machine Learning in Security	Production deployment in major security platforms	Validates AI-driven threat detection and alert prioritization
Federated Query Processing	Established in large-scale analytical systems	Enables cross-environment querying without centralized storage
Phased Implementation	Proven through incremental technology adoption	Allows organizations to validate effectiveness while managing change risk

6. Conclusions

The Federated Security Control Data Fabric represents a critical evolution in how enterprise security operations are architected for multi-cloud environments. Rather than attempting to centralize all security telemetry into monolithic SIEM platforms, the F-SCDF embraces the distributed reality of modern cloud deployments by keeping data close to its source while still enabling unified visibility and orchestration. This approach directly addresses the structural limitations of centralized architectures: unsustainable data egress and storage costs, processing delays that impede real-time detection, and fragmented visibility across provider-specific tools and schemas.

By combining distributed ingestion gateways, a universal security control schema with semantic mapping, federated lakehouse-style storage, and an AI-powered signal prioritization broker, the F-SCDF offers a cohesive pattern for managing preventive, detective, and remediative control telemetry at scale. Existing technologies—such as OpenTelemetry-style agents, lakehouse platforms like Apache Iceberg and Delta Lake, emerging semantic standards like OCSF, and mature machine learning techniques for alert triage—provide a solid technical foundation for practical implementation. A phased adoption strategy allows organizations to introduce these capabilities incrementally, validating cost savings, latency reductions, and analyst efficiency improvements without disruptive rip-and-replace migrations.

As cloud adoption deepens and telemetry volumes grow exponentially, the economic and operational shortcomings of SIEM-centric models will become increasingly pronounced, outpacing both security budgets and analyst capacity. Organizations that adopt federated, vendor-agnostic telemetry architectures early will gain structural advantages in detection effectiveness, cost efficiency, and architectural flexibility. The F-SCDF provides a sustainable blueprint for aligning security operations with the distributed nature of contemporary cloud ecosystems, enabling advanced threat detection and response without sacrificing economic or operational sustainability. Future work includes developing reference implementations across major cloud providers, defining benchmark datasets for multi-cloud telemetry, and quantitatively evaluating F-SCDF deployments against traditional centralized security architectures.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] Tanner Luxner, "Cloud computing trends: Flexera 2024 State of the Cloud Report," Flexera Software LLC, 2024. [Online]. Available: https://www.flexera.com/blog/finops/cloud-computing-trends-flexera-2024-state-of-the-cloud-report/

[2] Gartner, "Gartner Forecasts Worldwide Public Cloud End-User Spending to Reach Nearly $600 Billion in 2023," 2022. [Online]. Available: https://www.gartner.com/en/newsroom/press-releases/2022-10-31-gartner-forecasts-worldwide-public-cloud-end-user-spending-to-reach-nearly-600-billion-in-2023

[3] Zahra Shojaee Rad and Mostafa Ghobaei-Arani, "Federated serverless cloud approaches: A comprehensive review," Computers and Electrical Engineering, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0045790625003155

[4] Visalakshmi Suresh et al., "Scalable and responsive event processing in the cloud," PMC Journal, 2013. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC3538295/

[5] Hanane Chliah et al., "Performance Security in Distributed Systems: Comparative Study," ResearchGate, 2018. [Online]. Available: https://www.researchgate.net/publication/384632560_Performance_Security_in_Distributed_System_Comparative_Study

[6] Tarun Jain et al., "Procuring Cloud Services: An Economic Analysis of Multi-cloud Strategy," SAGE Publications, 2025. [Online]. Available: https://journals.sagepub.com/doi/10.1177/10591478251326421

[7] Piyush Patil, "Optimizing low latency public cloud systems: Strategies for network, compute, and storage efficiency," WJARR, 2025. [Online]. Available: https://journalwjarr.com/sites/default/files/fulltext_pdf/WJARR-2025-1538.pdf

[8] Boddepalli Jahnavi, "Mitigating Security Risks in Multi-Cloud and Hybrid Cloud Environments: Cross-Cloud Communication and Threat Detection Frameworks," IJSET, 2024. [Online]. Available: https://www.ijset.in/wp-content/uploads/IJSET_V13_issue3_153.pdf

[9] Aptin Babaei et al., "A Review of Machine Learning-based Security in Cloud Computing," ResearchGate, 2023. [Online]. Available: https://www.researchgate.net/publication/373838363_A_Review_of_Machine_Learning-based_Security_in_Cloud_Computing

[10] Zscaler, "Zscaler ThreatLabz 2025 VPN Risk Report with Cybersecurity Insiders". [Online]. Available: https://www.zscaler.com/campaign/threatlabz-vpn-risk-report