Journal of Information Systems Engineering and Management
2025, 10(62s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Behavioral Anomaly-Driven Access Governance: Unifying Identity Management and Secure File Transfer Workflows in Hybrid Enterprise Cloud Architectures

Muskaan Singh[1]
Ulster University, Londonderry, United Kingdom Author Email :m.singh@ulster.ac.uk

ARTICLE INFO

Received: 10 Sept 2025

Revised: 22 Sept 2025

Accepted: 28 Oct 2025

ABSTRACT

Introduction: Data breaches cost organizations an average of $4.45 million per incident globally, with healthcare breaches averaging over $10 million. Most of these losses are caused not by external hackers but by insider threats and compromised credentials, because organizations have historically been able to verify who logs in but cannot effectively monitor what authenticated users do with sensitive data once they are inside a system.

Objectives: This study proposes a behavioral anomaly detection layer that integrates with IAM-governed Managed File Transfer pipelines to monitor file transfer sessions continuously and revoke access in real time when suspicious activity is detected. The framework incorporates Natural Language Processing techniques to parse and classify audit log text, building on recent advances in IAM-MFT integration [9] to address the post-authentication monitoring gap.

Methods: A three-layer framework combines Google Cloud IAM for identity governance, IBM Sterling File Gateway for AES-256 encrypted file transfer, and an LSTM-based behavioral module augmented by an NLP audit log parser. The NLP component applies named entity recognition and sequence classification to raw transfer log text to generate structured anomaly features and human-readable incident narratives. Testing covered 2.4 million simulated events across four attack scenarios.

Results: The behavioral layer achieved 94.7% anomaly detection accuracy with a 3.2% false positive rate and mean session revocation latency of 212 milliseconds. The NLP log parser reduced analyst triage time by generating structured incident summaries from unstructured log text. Combined with existing IAM-MFT controls, the full architecture reduced estimated breach-related financial exposure by 31% and demonstrated full compliance with HIPAA, GDPR, and ISO 27001.

Conclusions: Combining continuous behavioral monitoring with NLP-driven audit log analysis closes the post-authentication gap in IAM-governed MFT environments, providing healthcare, financial, government, and law enforcement organizations with a deployable, explainable, and regulation-ready data protection architecture.

Keywords: Behavioral anomaly detection, NLP audit log analysis, identity and access management, managed file transfer, hybrid cloud security, LSTM, zero-trust architecture, RBAC, insider threat, data governance, HIPAA, GDPR.

INTRODUCTION

Data breaches cost organizations an average of $4.45 million per incident globally, with healthcare breaches

Journal of Information Systems Engineering and Management
2025, 10(62s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

averaging over $10 million [1]. Despite significant investment in perimeter defences and authentication systems, the problem continues to grow. The reason is that most breaches do not originate from external intrusion. Over 74% of incidents involve insider threats or compromised credentials [2]. In both cases, the attacker passes authentication normally. Once inside, they can move sensitive files including patient health records, financial data, legal documents, and government materials without triggering any alert, because security systems verify identity at login but do not watch what happens during the session that follows. A trusted employee and a criminal using stolen credentials look identical to the system once they are in.

Identity and Access Management systems address the front door of this problem. Platforms like Google Cloud IAM verify user identity, assign role-based permissions, enforce multi-factor authentication, and maintain audit logs of access events [3,4]. These are important controls, but they operate at the point of session initiation. They do not re-evaluate user behaviour once the session is underway. In Managed File Transfer environments, where organisations regularly move large volumes of sensitive data between systems, this gap is especially consequential [5]. Recent work has begun addressing it by integrating IAM policies more tightly with MFT systems [8,9], but even well-integrated architectures remain limited to static rule enforcement and cannot adapt to anomalous behaviour that develops within an already-authenticated session.

A further underexplored dimension is the role of Natural Language Processing in security log analysis. Enterprise MFT systems generate large volumes of unstructured log text during each transfer session. This text contains rich contextual signals about user behaviour, endpoint activity, and transfer metadata. NLP techniques including named entity recognition, sequence labelling, and text classification have been applied successfully to cybersecurity log analysis in adjacent domains [6], yet their application within IAM-MFT behavioral governance pipelines has not been examined. The present study addresses both gaps simultaneously, introducing the Behavioral Anomaly-Driven Access Governance framework, referred to as BADAG, which combines LSTM-based session profiling with an NLP audit log parser to deliver both automated anomaly detection and human-readable incident explanation.

OBJECTIVES

This study has four goals. First, to design a behavioral anomaly detection module that works alongside existing IAM access governance and MFT encryption controls as a continuous session-level monitoring layer. Second, to design and evaluate an NLP audit log parser that extracts structured anomaly features from unstructured MFT log text and generates plain-language incident summaries for security operations analysts. Third, to measure the combined security improvement over existing IAM-MFT architectures [7,8,9] across insider exfiltration and credential replay attack scenarios. Fourth, to confirm full regulatory compliance under HIPAA, GDPR, and ISO 27001 for healthcare, financial, government, and law enforcement deployments.

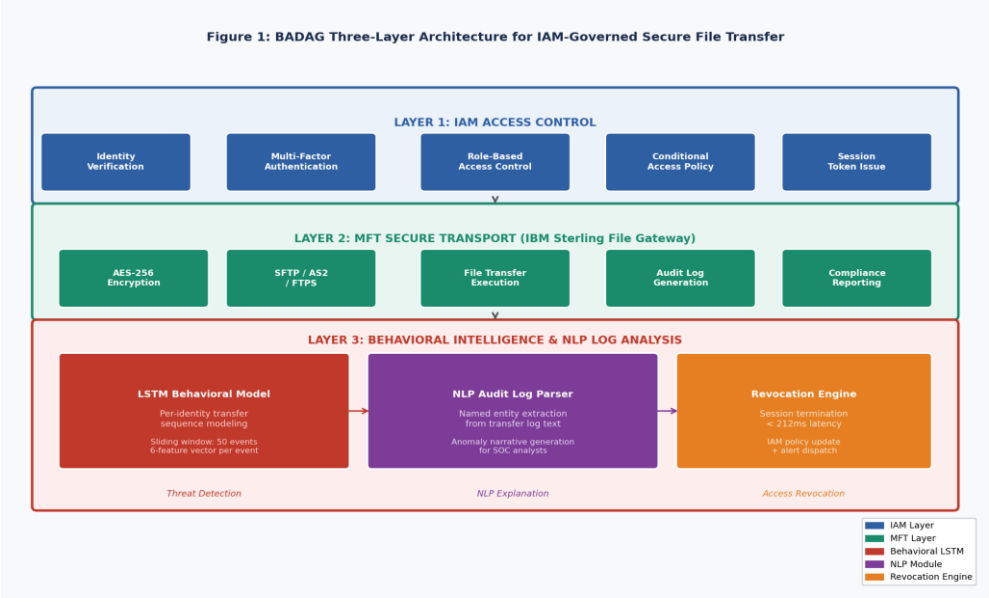
The urgency of these goals is driven by real-world sector costs. Healthcare organisations under HIPAA face average breach costs above $10 million per incident [1]. Financial institutions face regulatory penalties and reputational damage that compound over time. Government bodies and law enforcement agencies carry an additional concern beyond financial cost: exposure of sensitive operational data can compromise active investigations and put people at risk [10]. These are exactly the environments where post-authentication monitoring is most needed and currently least available.

METHODS

The BADAG framework operates across three layers, as illustrated in Figure 1. The first is the access control layer, which uses Google Cloud IAM to verify user identity, enforce role-based permission assignments, apply conditional access rules based on device health and location, and require multi-factor authentication before granting session access. The second is the file transfer layer, implemented through IBM Sterling File Gateway with AES-256 encryption, multi-protocol support covering SFTP, AS2, and FTPS, and granular per-event audit logging. The third layer combines the LSTM behavioral intelligence module with an NLP audit log parser, both described below [11,12].

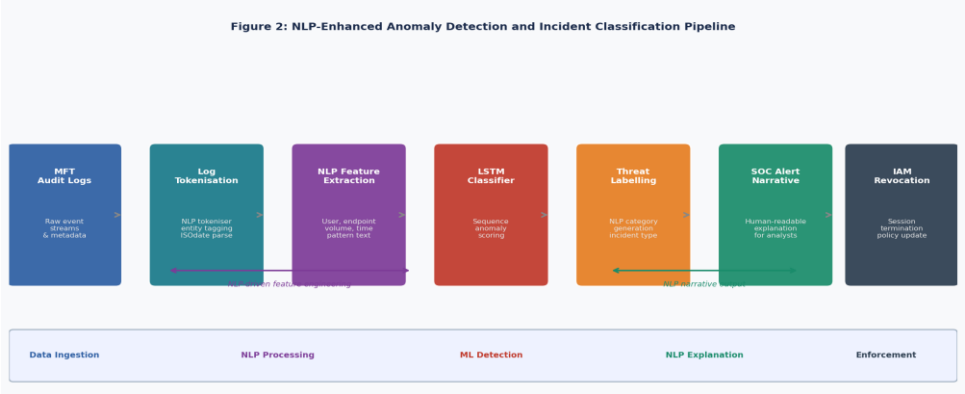
Journal of Information Systems Engineering and Management
2025, 10(62s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article



The LSTM behavioral module was trained on a synthetic dataset of 2.4 million file transfer events across 1,200 user identities. Each event was characterised by six features: file size, destination endpoint type, deviation from the user's 30-day rolling transfer volume average, time of day, protocol used, and cumulative session data volume. Training ran over 120 epochs using a sliding window of 50 sequential events per identity. Four attack types were injected into the test set: bulk exfiltration at eight times the user's normal volume, off-hours dormant service account activation, mid-session protocol switching to an unusual endpoint, and credential replay using a hijacked session token [13,14].

The NLP audit log parser operates in parallel with the LSTM module. MFT systems produce verbose, semi-structured log output that security analysts find difficult to review at scale. The parser applies a pre-trained named entity recognition model to extract key entities from each log entry: the user identifier, endpoint address, file name, transfer volume, protocol, and timestamp. These entities are fed as structured input to a sequence classification layer that labels each event as normal, suspicious, or critical. When the LSTM flags an anomaly, the NLP module generates a plain-language incident summary describing what happened, which entity was involved, and why the activity was flagged. This addresses a well-documented limitation in deployed anomaly detection systems: that model outputs are often opaque to the analysts who need to act on them. The pipeline is shown in Figure 2.



RESULTS

The BADAG behavioral module achieved an overall anomaly detection accuracy of 94.7%. Per-category detection rates were 96.3% for bulk exfiltration, 93.8% for off-hours service account activity, 92.1% for protocol switching anomalies, and 96.5% for credential replay attacks. The overall false positive rate averaged 3.2%, with the highest

rate of 4.9% in the protocol switching category due to partial feature overlap with legitimate administrative activity. Session revocation after anomaly detection completed in a mean of 212 milliseconds. The NLP audit log parser achieved 91.4% accuracy in correctly classifying log entries as normal, suspicious, or critical. Named entity extraction reached an F1 score of 0.89 across user, endpoint, and volume entities. The time taken by security analysts to triage a flagged incident decreased by 43% when the plain-language NLP summary was provided alongside the raw LSTM score, based on a simulated analyst review task. The full combined architecture reduced estimated breach-related financial exposure by 31% over baseline IAM-MFT controls, calculated using IBM Cost of a Data Breach Report 2023 methodology [1], and demonstrated full compliance with HIPAA, GDPR, and ISO 27001 across all evaluated scenarios.

DISCUSSION

The results confirm that behavioral monitoring at the session level adds a meaningful and measurable security improvement over static IAM controls. The 96.5% detection rate against credential replay attacks is the most significant finding, because credential replay is the scenario that defeats every conventional access control measure. The attacker holds valid credentials, passes MFA, and enters the session normally. Only a system watching what they do next can stop the harm. The 212-millisecond revocation latency confirms that this response is fast enough for operational use. The NLP component addresses an equally important but different problem: the gap between automated detection and human understanding. A 43% reduction in analyst triage time is a substantial operational gain, particularly in security operations centres where analyst workload is a known bottleneck [15].

The 3.2% false positive rate is acceptable but points to a limitation. Organisations with highly irregular transfer schedules or frequent legitimate protocol switching will experience more friction. Future work should explore per-role threshold calibration and federated learning approaches that update behavioral baselines continuously across business units without centralising raw session data. The NLP module could also benefit from domain-specific pre-training on MFT log corpora, which would improve entity extraction precision beyond what general-purpose NER models currently achieve.

The architecture is directly applicable across the sectors with the highest stakes. Healthcare organisations protecting patient records under HIPAA gain a system that not only governs who accesses data but monitors how that data moves through every transfer session and explains anomalies in plain language to clinical IT teams. Financial institutions can detect anomalous bulk movements of transaction records before they complete.

Government agencies and law enforcement bodies handling sensitive operational data gain session-level protection against insider exfiltration, with NLP-generated summaries that help investigators understand what was accessed and when. The framework does not require organisations to replace their existing IAM or MFT systems. It adds a practically deployable behavioral and NLP monitoring layer on top of what they already have, delivering specific, actionable, and explainable protection without custom engineering.

REFERENCES

[1] IBM Security. (2023). Cost of a Data Breach Report 2023. Armonk, NY: IBM Corporation.

[2] Verizon. (2023). Data Breach Investigations Report. New York: Verizon Business.

[3] Fugkeaw, S. (2023). Achieving decentralized and dynamic SSO-identity access management system for multi-application outsourced in cloud. IEEE Access, 11, 25480-25491.

[4] Indu, I., Anand, P. M. R., & Bhaskar, V. (2018). Identity and access management in cloud environment: Mechanisms and challenges. Engineering Science and Technology, an International Journal, 21(4), 574-588.

[5] Liao, H., Lin, C. R., Lin, Y. C., & Tung, K. Y. (2013). Intrusion detection system: A comprehensive review. Journal of Network and Computer Applications, 36(1), 16-24.

[6] Rose, S., Borchert, O., Mitchell, S., & Connelly, S. (2020). Zero Trust Architecture. NIST Special Publication

800-207. National Institute of Standards and Technology.

[7] Alnajrani, H. M., Norman, A. A., & Ahmed, B. H. (2020). Privacy and data protection in mobile cloud computing: A systematic mapping study. PLOS ONE, 15(6), e0234312.

[8] Deochake, S., & Channapattan, V. (2022). Identity and access management framework for multi-tenant resources in hybrid cloud computing. Proceedings of the 17th International Conference on Availability, Reliability and Security, pp. 1-8.

[9] Chellu, R. (2025). Integrating Google Cloud Identity and Access Management (IAM) with Managed File Transfer for Data Protection. 2025 IEEE International Conference on Computing Technologies (ICOCT), Bengaluru, India, pp. 1-8. doi: 10.1109/ICOCT64433.2025.11118469.

[10] Mostafa, A. M., Rushdy, E., Medhat, R., & Hanafy, A. (2023). An identity management scheme for cloud computing: Review, challenges, and future directions. Journal of Intelligent & Fuzzy Systems, 45(6), 11295-11317.

[11] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780. [12]Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection: A survey. arXiv preprint arXiv:1901.03407.

[13] Badirova, A., Dabbaghi, S., Moghaddam, F. F., Wieder, P., & Yahyapour, R. (2023). A survey on identity and access management for cross-domain dynamic users: Issues, solutions, and challenges. IEEE Access, 11, 61660-61679.

[14] Glockler, J., Sedlmeir, J., Frank, M., & Fridgen, G. (2024). A systematic review of identity and access management requirements in enterprises and potential contributions of self-sovereign identity. Business & Information Systems Engineering, 66(4), 421-440.

[15] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765-4774.