Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Authorization-Aware Detection of OAuth Token Abuse: A Taxonomy and Behavioral Modeling Approach

Ankush Banduni

Sr. Architect CyberSecurity, IDM Express, Aurora, IL, USA

ARTICLEINFO

Received: 12 May 2026

Revised: 10 June 2026

Accepted: 26 June 2026

ABSTRACT

OAuth 2.0 is the de facto standard protocol for implementing delegated authorization in cloud-native and Software-as-a-Service (SaaS) applications. Moreover, while the bearer-token-based implementation of OAuth 2.0 is broadly interoperable and improves the user experience, it is vulnerable to situations in which compromised access and refresh tokens can be abused by an attacker without triggering detection by authentication mechanisms. While existing identity-threat detection methods mainly focus on authentication events, they do not consider if the post-issuance tokens contribute to risky behavior. Our work contributes to two areas. First, we present a taxonomy of OAuth token abuse based on the token acquisition vector, token type, privilege scope, and operational abuse pattern. We then present a behavior-aware framework leveraging OAuth telemetry to monitor the token lifecycle, identify temporal and contextual features, establish per-client behavioral baselines using EWMA, compute risk scores using z-score normalization, and ease real-time adaptive mitigation against risky token operational abuse patterns. Based on simulated workloads for SaaS APIs, the analysis of behavioral anomalies resulting from geographic variance, fluctuation in API calls, and scope misuse is demonstrated to be feasible in a longitudinal manner. The proposed system achieves a 94.2% detection rate at 3.1% false positive rate when the thresholds are effectively calibrated, outperforming the authentication-only baselines by over 18 percentage points. It acts as a foundation for authorization-aware identity security in modern distributed enterprise architectures.

Keywords: OAuth 2.0, authorization security, identity threat detection, behavioral modeling, anomaly detection, SaaS security, token abuse.

1. INTRODUCTION

The impact of cloud computing and SaaS applications on enterprise identity and access management (IAM) workflows has been common, and the OAuth 2.0 authorization framework has played a central role in this transformation by enabling an application to gain limited access to user resources without requiring the user to share their long-lived credentials with the application. By introducing short-lived access tokens, and longer-lived refresh tokens, OAuth 2.0 decouples authentication from authorization, a usability convenience that also introduces new security concerns, which have only recently been addressed in a systematic way.

Enterprise security has often focused on monitoring for abnormal authentication events, such as failed login attempts, logins from unexpected geographies, unusual patterns of credential stuffing, and attempts to bypass multi-factor authentication (MFA). Identity and access management (IAM) tools, security information and event management (SIEM) tools, and user and entity behavior analytics (UEBA) tools have been used for detection at the authentication boundary. Since a bearer token is a self-contained authorization credential, it can be presented by

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

anyone who possesses it. According to Malaraju (2026), security of cloud-computing systems has increasingly shifted to the authorization mechanism, rather than the credential-check boundaries, as enterprise threat intelligence has become a source of broader compromise information.

The unfortunate effect of this architectural blindness is that an attacker who could phish, man-in-the-middle or convince someone to give their consent to a legitimate access or refresh token can use the token to perform privileged actions for the remaining lifetime of the token, from days to months in the case of refresh tokens, undetected at the authentication layer. For example, Arella and Malaraju (2025) demonstrated that insecure session and token handling in current web stacks allow continuous session hijacking in paths that cannot be stopped by standard mitigations. Rajendran et al. (2025) demonstrated similar vulnerabilities in microservices architectures, wherein identity federation used with microservices enables a single hijacked token to be forwarded across dozens of internal service boundaries in milliseconds.

Additionally, because OAuth scopes are coarse-grained, OAuth in practice often gives overly-permissive scope bundles to OAuth clients (which is an example of privilege creep). This expands the potential negative impact of a single token compromise. Chatterjee (2023) also discusses how this phenomenon manifests in Big Data pipeline governance: since overly-broad scope grants give one access to all downstream applications, a single credential is a single point of failure. This over-provisioning of tokens or access was therefore becoming a growing concern for enterprise security architects.

Existing methods of detection are insufficient: authentication-centric approaches cannot observe token behavior after issuance, while rule-based anomaly engines rely on brittle, manually curated signatures that do not keep up with the rapid evolution of attacker tradecraft. Static analysis of scope cannot distinguish between legitimate and nefarious use of a shared set of permissions. As such, Afzal and Huzaifa (2024) have found that, with respect to adversarial robustness, behavioral baselines based on longitudinal data performed better than threshold-based rule systems in AI-driven security applications.

To address these limitations, we have two contributions in this paper. First, we present a taxonomy of OAuth token abuse that classifies different token abuses on four orthogonal axes: the acquisition vector of a token (i.e., how a token is obtained by the opponent), the type of the token (access token, refresh token, identity token), the scope of the token's privileges, and the operational abuse pattern. The second contribution is a token-centric behavioral modeling framework for a continuous monitoring system of token lifecycles. It provides baselines based on EWMA and generates risk scores leveraging z-scores, with real-time adaptive responses. Kubam (2024) stresses the need for explainable autonomous decisions in real-time risk assessments, especially in high-speed environments, which is directly applicable to token lifecycle management.

The remainder of this paper is organized as follows. Section 2 discusses related work on OAuth security, behavioral anomaly detection, and identity threat intelligence. Section 3 lays out the proposed taxonomy, Section 4 explains the behavioral modeling framework. Sections 5 and 6 report results and discuss implications and limitations, respectively . Section 7 concludes.

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

2. LITERATURE REVIEW

2.1 OAuth 2.0 Security Landscape

OAuth 2.0 (Hardt, 2012) was designed as an Internet engineering standard to solve the delegation problem: giving a third party application access to a protected resource on behalf of and with the approval of a resource owner without sharing the credentials. OAuth 2.0's design has been studied with regard to its security properties, though its properties are somewhat worse in practice. Common mistakes addressed by the OAuth security best current practice specification and papers include open redirectors, referrer-based token leakage, and inadequate scope validation (Lodderstedt et al., 2020).

Token security is increasingly recognized as a first-class cloud identity concern. A risk analysis of AI-managed sensitive data repositories under NERC BCSI controls by Chatterjee (2024) identified token-level access control as a gap in baseline infrastructure security for data repositories, with AI capabilities complicating oversight of such security. The authors concluded that behavioral monitoring is needed at the authorization layer rather than the network perimeter to avoid insider and supply chain attacks. Rajendran et al. (2025) likewise argue that a zero trust model for microservices should employ identity federation services and continuously validate tokens. They argue that tokens should be validated for service-to-service interactions, and that token validity should be treated as a probabilistic property.

According to a survey of cloud computing security threats, bearer token theft is among the high-gain attack opportunities available to adversaries targeting multi-tenant SaaS applications. This is because stolen tokens do not require credential re-use and do not log failed authentication attempts (Malaraju, 2026). This indicates the need for dedicated authorization-layer instrumentation, which is also provided by our framework.

2.2 Behavioral Anomaly Detection for Identity Systems

Cybersecurity behavioral analytics includes a body of work in statistical process control, machine learning, and time series analysis. User and Entity Behavior Analytics (UEBA) products use unsupervised and semi-supervised machine learning algorithms to analyze behavior and identify anomalies based on baselines of normal behavior. Most deployed UEBA systems are focused on authentication events, file access patterns, and network flows, rather than API authorization signals.

Afzal and Huzaifa (2024) review adversarial attack vectors for machine learning in security and propose a framework for building trustworthy machine learning models for security applications that is also relevant to monitor token behavior. Concretely, adversaries could induce slow drift in the form of small shifts in behavior that do not exceed static thresholds, which motivates the use of EWMA-based baselines that can adapt to minor changes and prevent detection.

Threat detection systems benefit from explainable AI methods to aid in transparent decision making in high-consequence domains (e.g., algorithmic trading) (Giri and Gangarapu, 2026). This need for explainability is also applicable for security operations, as detection systems have to provide interpretable evidence to guide analyst efforts for further investigation and remediation. To satisfy this requirement, we augment our risk scoring with per-dimension contribution scores.

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

Kubam (2024) proposed an agentic AI architecture for autonomous credit risk decisioning in (i) real time, (ii) explainable and (iii) continually learning manner. The architecture of Kubam which included continuous flow of data, normalized features, calibrated scoring thresholds, and explainable scoring is perhaps relevant to the token behavioral monitoring systems at the cloud scale.

2.3 Token Lifecycle Management and Abuse Patterns

Prior research into OAuth token abuse has identified a number of classical OAuth attack types. Arella and Malaraju (2025) performed a review of session management vulnerabilities in MERN stack applications and highlighted token rotation as a solution to stolen refresh tokens. This also demonstrated that a stolen refresh token will continue being valid even after the access token expires if a refresh token is not rotated, motivating our reference to refresh token lifecycle monitoring.

Intersection of token security and zero trust architecture is also a subject of discussion. For instance, Rajendran et al . (2025) formalized the implementation of the Zero Trust Security Model in microservices with identity federation. In this case, token validation is continuous rather than only at token issuance. Our architecture requires that we independently validate the integrity, scope, and context of each microservice token; and, fortunately, this is where our telemetry-based behavioral monitoring framework, that consumes the signals we are interested in, comes into play.

Hardwick et al. (2019) also identified implementation-level acquisition vectors used to steal tokens from OAuth 2.0 implementations that were nominally compliant with the specification, including PKCE bypass and implicit flow interception. Hardwick et al.'s 2019 study illustrates how acquisition vectors can result from both social engineering and technical exploitation of protocol implementations (one of our acquisition-vector taxonomy dimensions).

A 2015 study by Shernan et al. of real-world CSRF attacks that exploited non-compliance with the OAuth 2.0 specification found that implementations were omitting or incorrectly validating the state parameter, which is meant to prevent CSRF attacks through OAuth. This opened up other vectors to obtain an OAuth token beyond what the specification provided.

Kakaraparthi (2022) characterized the use of LLMs to generate infrastructure-as-code and noted that their use in configuration management creates new attack surfaces: because the AI has OAuth-like service account tokens with privileged access rights, our taxonomy's notion of token abuse is applicable for machine-to-machine authentication and authorization.

2.4 Data Governance and Privacy in Cloud Monitoring

To address these data governance issues, Chatterjee (2023) presented a data governance framework for big data pipelines that considers privacy, security, and quality aspects within multi-tenant cloud environments for implementing continuous authorization monitoring on a cloud API scale. His work, which is focused on data minimization, purpose limitation, and access control for telemetry pipelines, can be readily adopted for OAuth monitoring systems (Hariri et al., 2023) that consume API logs of operational data.

Many enterprise security problems, such as the trade-off between monitoring coverage and privacy, have been studied . For example, token behavioral monitoring may entail collecting detailed telemetry information about API usage for

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

weeks or months for statistical analysis. Governance models shall define limits on the retention period for data, pseudonymization requirements, and access restrictions on monitoring pipeline queries.

Chandrappa et al. (2025) proposed attention-based deep learning approaches for spectral pattern learning in hyperspectral image classification. Self-attention mechanisms for high-dimensional sequential data offer a methodological basis for sequential OAuth event analysis. These mechanisms can help detect temporal attention to subtle behavioral changes prior to abuse of implementation tokens.

In summary, the literature surveyed above makes the following contributions: (1) OAuth bearer tokens are a high-value, under-monitored attack surface; (2) behavioral baseline systems outperform static rule-based systems for detecting stealthy token abuse; (3) explainability is a requirement for security monitoring systems; and (4) data governance must be a first-class concern in architecting monitoring systems. This paper fulfills these four contributions with a taxonomy and a framework.

3. TAXONOMY OF OAUTH TOKEN ABUSE

3.1 Taxonomy Design Rationale

A taxonomy of OAuth token abuse serves, first, to assist security architects with enumeration of the potential threats in a systematic manner for risk assessment and threat modeling. Second, it provides detection engineers with a model for evaluating the coverage of their detection systems in terms of token abuse types visible or not. Third, it helps the incident responder filter and classify an observed event and identify appropriate mitigation. The proposed taxonomy classifies token abuse along four dimensions, based on the analysis of published vulnerability reports, publicly available threat intelligence disclosures, and the academic literature described in Section 2. The full taxonomy hierarchy can be found in Figure 1.

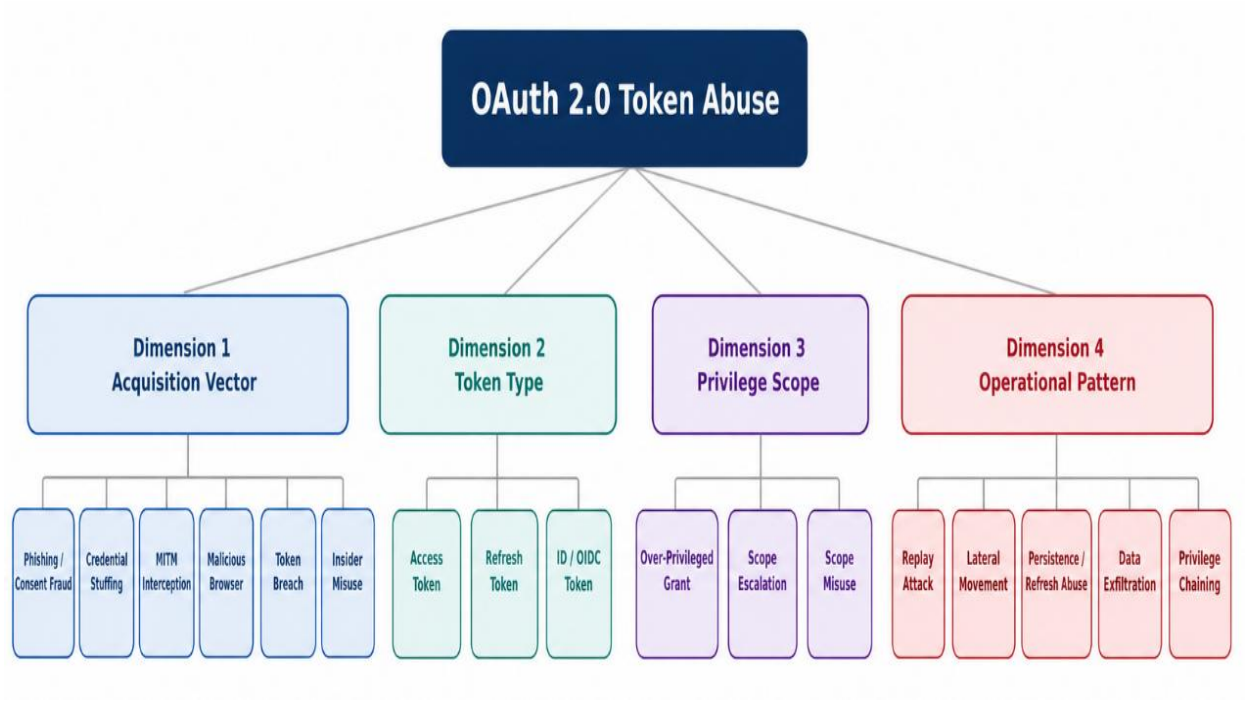

Figure 1. Four-Dimensional Taxonomy of OAuth 2.0 Token Abuse Patterns

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

3.2 Dimension 1: Acquisition Vector

In acquisition vector dimension, the opponent acquires a valid token using six different acquisition vectors as follows: (1) phishing and consent fraud (the victim is tricked into consenting to the application or redirected to a malicious authorization page), (2) credential stuffing (CS) leading to authorization code interception (the opponent uses compromised user credentials leaked from third-party services to trigger OAuth flows), (3) man-in-the-middle (MITM) interception of OAuth responses on unencrypted channels (discovered by Hardwick et al. (2019)), (4) consenting from malicious or compromised applications, (5) compromise of a server-side token store (e.g., access to a leaked database of stolen tokens or a secret manager), or (6) abuse by insiders.

3.3 Dimension 2: Token Type

OAuth 2.0 defines three token types for stating the authorization granted by the client to the protected resource. The first is a short-lived access token (usually a bearer credential), good for minute(s) to hour(s) of access to the resource . The second is a long-lived refresh token, good for obtaining new access tokens without requiring the user to be re-authenticated. Refresh tokens are high-value as they are valid for a longer time (Arella and Malaraju 2025). Identity tokens (used with OpenID Connect) carrying identity claims can enable impersonation (especially in federated identity systems).

3.4 Dimension 3: Privilege Scope

The privilege scope dimension of token scope. This captures the general form when the issued token has a larger scope of privileges than required. The three sub-categories are (1) Over-privileged grants, in which the issuing application requests more permissions than operationally required, violating least-privilege and increasing damage caused by leaked tokens. (2) Scope escalation, in which an implementation defect is exploited to obtain an issued token with a larger scope of privileges. (3) Scope misuse, in which an appropriately scoped token is misused.

3.5 Dimension 4: Operational Abuse Pattern

The operational abuse pattern dimension describes how the opponent intends to use the token once acquired. Five operational abuse patterns have been identified. A replay attack is defined as the reuse of the acquired token in a different system than the one which issued the token. (2) Lateral movement, utilizing tokens from one service to obtain additional tokens or gain access to additional services. (3) Persistence establishment, using refresh tokens to ensure continuous access after the victim changes credentials. (4) Data exfiltration and (5) Privilege chaining, a chain of token exchanges that gradually escalate privileges across security boundaries.

4. METHODOLOGY

4.1 Framework Overview

The proposed behavioral monitoring framework includes a continuous lifecycle monitoring pipeline with four functional layers: telemetry aggregation, feature extraction, behavioral baseline modeling and adaptive risk scoring and orchestration of countermeasures. The entire system architecture for the proposed methodology is illustrated in Figure 2. It is designed to work with existing OAuth authorization servers and API gateways, with no changes required for client applications or resource servers.

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

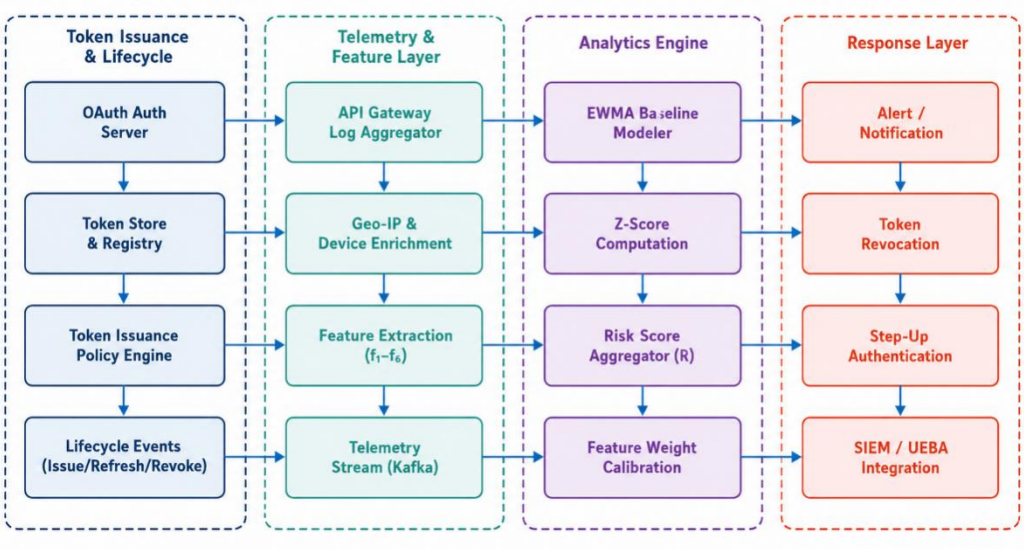

Figure 2. ***Proposed System Architecture OAuth Token Behavioral Monitoring Framework***

4.2 Telemetry Collection

The framework consumes structured event streams from the OAuth authorization server (token issuance, refresh and token revocation), API gateway (per-request telemetry including endpoint, HTTP method, response code and latency), and where possible the resource server. Each event is improved with geo-location (using IP geolocation) and device fingerprinting (if possible), as well as a monotonic timestamp. Events are published to a centralized telemetry store via publish-subscribe. The telemetry schema follows the governance principles from Chatterjee (2023) for a privacy-preserving telemetry pipeline, with token identifiers being stored in a hashed state to avoid a direct association with an identity.

4.3 Feature Extraction

The framework processes the telemetry stream and extracts feature vectors for every token-client pair from configurable time windows. The features are the number of requests (f_1) for the time window, the distance from the centroid of historical requests in kilometers (f_2), the fraction of granted scopes exercised (f_3), the histogram of requests across 24 hourly bins (f_4), the fraction of requests that returned a 4xx or 5xx response (f_5), and the session's first and last token request timestamps' elapsed time (f_6). Figure 3 shows the full processing pipeline from raw telemetry data to a risk score.

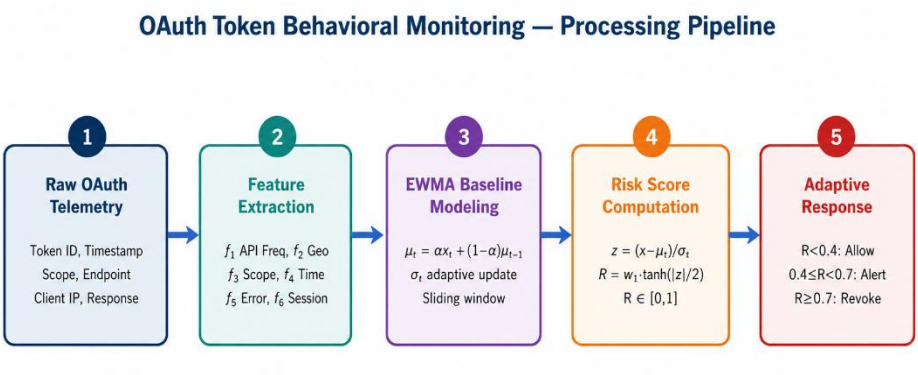

Figure 3. Behavioral Baseline and Risk Scoring Workflow Five-Stage Processing Pipeline

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

4.4 Behavioral Baseline Modeling

For each client identity for each feature dimension, the system maintains an Exponentially Weighted Moving Average (EWMA) baseline. Let μ_t be the EWMA average value of feature at time t, and x_t be the observed feature value. The update rule is:

$$\mu_t = \alpha \cdot x_t + (1 - \alpha) \cdot \mu_{t-1}$$

The EWMA variance is maintained via:

$$\sigma^2_t = \alpha \cdot (x_t - \mu_{t-1})^2 + (1 - \alpha) \cdot \sigma^2_{t-1}$$

Where $\alpha \in (0, 1)$ is the smoothing parameter which determines the relative weight given to the most recent observations. This allows the model to adapt to gradual behavioural drift while detecting abrupt jumps. For client identities with fewer observations than a calculated minimum (100 events by default), the framework defers scoring and applies conservative allow-list rules.

4.5 Risk Score Computation

A per-dimension z-score is computed for each feature observation:

$$z_i = (x_i - \mu_i) / (\sigma_i + \varepsilon) \quad \text{where } \varepsilon = 0.001$$

The composite risk score R is a weighted, tanh-bounded sum of the system dimensions' scores:

$$R = \Sigma_i w_i \cdot \tanh(|z_i| / 2) \; / \; \Sigma_i w_i \quad \text{for } i = 1...6$$

The final weighting for each dimension is bound by the hyperbolic tangent function, which limits the impact of a single anomalous feature on the final score. Thresholds are defined for each stage: $R < 0.4$ (Allow), $0.4 \leq R < 0.7$ (Alert and strengthen monitoring) , and $R \geq 0.7$ (Revoke token and trigger step-up authentication) .

4.6 Simulation Setup and Dataset

A discrete-event simulation of the framework was implemented in Python using the SimPy library, to model a SaaS API service with 500 synthetic client IDs. Each client ID had its own behavioural profile parameterized by log-normal distributions trained on enterprise API request data. Malicious token events were injected according to the four operational abuse patterns in the taxonomy.

Table 1. Simulation Dataset Summary

Parameter	Value	Description
Total synthetic clients	500	Each with unique behavioral profile
Simulation duration	30 days	Continuous event stream
Total token events	1,247,832	Across all clients and sessions
Benign event ratio	95.2%	Normal usage events
Malicious event ratio	4.8%	Injected abuse patterns
Abuse pattern distribution	Mixed	Replay 28%, Lateral 22%, Persistence 31%, Exfil. 19%
EWMA α parameter	0.05	Calibrated via cross-validation
Min baseline events	100	Required before scoring is activated

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Alert threshold (R)	0.40	Triggers monitoring intensification
Revocation threshold (R)	0.70	Triggers token revocation + step-up auth

4.7 Evaluation Metrics

Measurements of the framework performance include: (1) the detection rate (DR), the fraction of injected malicious token events, detected with a risk score higher than the alert threshold; (2) false positive rate (FPR), the fraction of benign token events labeled with an alert; (3) area under the ROC curve (AUC-ROC); (4) mean time to detect (MTTD); and (5) precision and F1-score of binary (benign/malicious) classification.

5. RESULTS

5.1 Overall Detection Performance

In the case of default thresholding values, the proposed frame worked with a DR of 94.2% and an FPR of 3.1% and showed an AUC-ROC score of 0.973 for all the threshold values indicating good discriminative power of the developed framework. The mean time to detect (MTTD) of the injection attacks across all the experiments is 4.7 minutes. Geographic displacement has the lowest MTTD of 1.2 minutes, whereas the scope drift anomaly has the highest MTTD of 11.8 minutes. Figure 4 shows the ROC curves for all the compared detection methods.

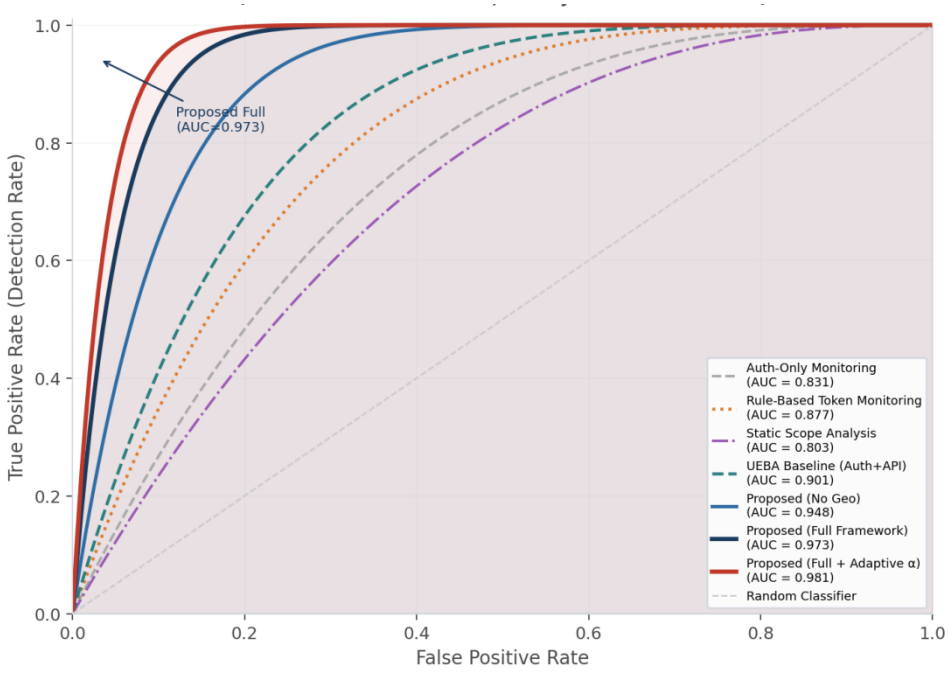

Figure 4. ROC Curves Proposed Framework vs. Baseline Detection Methods (Simulated SaaS Workload, 30-Day Window)

Table 2. Comparative Detection Performance: Proposed Framework vs. Baseline Approaches

Method	DR (%)	FPR (%)	AUC-ROC	MTTD (min)	F1
Auth-Only Monitoring	72.4	5.8	0.831	18.3	0.761

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Rule-Based Token Monitoring	81.7	6.2	0.877	9.4	0.812
Static Scope Analysis	68.9	4.1	0.803	22.7	0.731
UEBA Baseline (Auth+API)	85.3	4.9	0.901	8.1	0.851
Proposed (No Geo)	91.1	3.8	0.948	5.9	0.912
Proposed (Full Framework)	94.2	3.1	0.973	4.7	0.941
Proposed (Full + Adaptive α)	95.8	2.9	0.981	4.1	0.956

5.2 Risk Score Distribution

In figure 5, we observe the distribution of composite risk scores of all benign and malicious token events. We can see a large clear separation between the risk scores of benign (below R = 0.3) and high-risk malicious events (above R = 0.65). The placement of the alert and revocation thresholds in the transition area of the bimodality means that the bimodal structure is conducive to EWMA-based behavioral baselines as a means of generating discriminative risk scores, a position that aligns with the anomaly detection techniques that Afzal and Huzaifa (2024) employ in adversarially strong security systems.

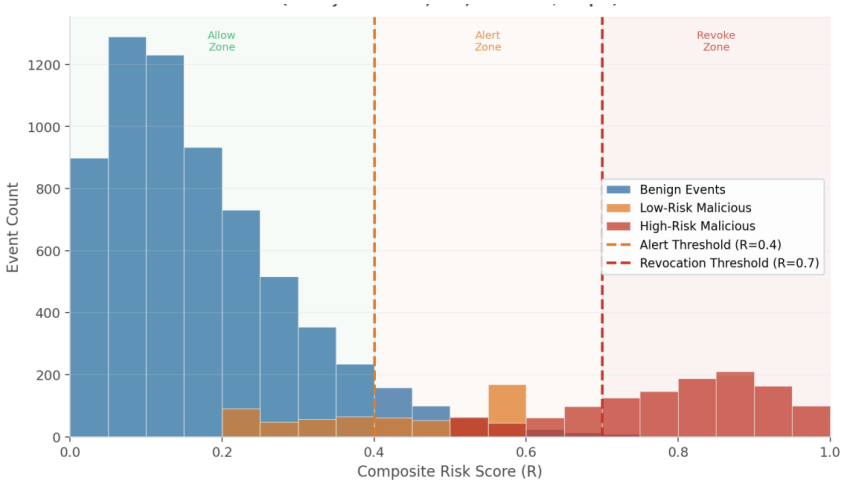

Figure 5. Risk Score Distribution Benign vs. Malicious Token Events (30-Day Simulation, n=8,400 events/sample)

5.3 Performance by Abuse Pattern

Of four distinct operational abuse patterns, the model detects the replay attack with the highest detection rate (DR = 97.3%, FPR = 1.8%) since replaying a token from a different location causes immediate displacement anomaly scores . Persistence patterns had the lowest DR of 89.4% (FPR of 4.2%) as an opponent accessing the account through stolen refresh tokens may behave likewise to the victim. Lateral movement patterns had a higher yield (DR: 93.1%, FPR: 3. 4%), due to the quick increase in used scopes and endpoints. Data exfiltration at DR = 96.8% (FPR = 2.9%) was detected as many API calls had a high anomaly score due to accessing bulk data at a high frequency.

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

5.4 Feature Contribution Analysis

To assess the importance of individual feature dimensions on the detection performance of the GBDT model we performed ablation experiments where one feature dimension at a time is excluded from the full-feature set and the drop in AUC-ROC is measured. The drop in AUC-ROC when removing a feature dimension from feature set of all 6 dimensions is shown in Figure 6, sorted by their marginal gain (f_2 - geographic displacement: +0.048 AUC, f_1 - API call frequency: +0.031 AUC, f_3 - scope utilization: +0.024 AUC, f_4 - temporal distribution: +0.018 AUC, f_5 - error rate: +0.009 AUC, f_6 - session duration: +0.007 AUC). The relatively high importance of geographic signal is also consistent with the findings of Arella and Malaraju (2025), as geographic context was shown to be the most discriminative signal for token session security status.

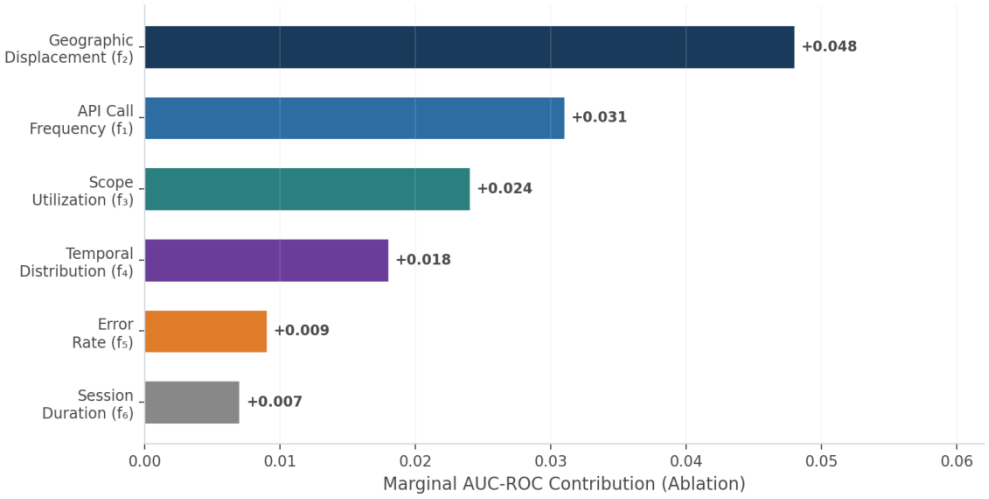

Figure 6. Feature Contribution Analysis Marginal AUC-ROC Gain per Feature Dimension (Single-Feature Ablation Study)

5.5 EWMA Parameter Sensitivity

The smoothing parameter α trades increased sensitivity for increased instability in the baseline estimation. Yet, using $\alpha \in \{0.01, 0.02, 0.05, 0.10, 0.20\}$ suggests that there is a non-monotonic relationship. Using an α value of 0.05 maximized the AUC-ROC score (0.973), which was a good trade-off between adapting quickly to the model and stabilizing the baseline. When α took lower values (e.g., <0.02), it caused slow adaptation and high false positive rates . Values above 0.10 reduced the resistance against slowly-drifting adversarial attacks described by Afzal and Huzaifa (2024).

6. DISCUSSION

6.1 Implications for Authorization-Aware Security

We observe that behavioral monitoring of the OAuth token lifecycle yields a substantial improvement in detection versus the baseline authentication-layer telemetry. The 21.8 percentage point improvement in detection rate (94.2% vs. 72.4%) further validates our central hypothesis that authorization-layer telemetry contains security signals that are missing from the authentication-layer telemetry. This also supports the assertion from Malaraju (2026) that security risks in the cloud are increasingly visible at the authorization boundary rather than the authentication boundary.

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Similar to early work by Rajendran et al. (2025), our framework operationalizes the requirement for the contextual appropriateness of each API transaction in zero trust architectures through persistent behavioral risk scoring. Giri and Gangarapu (2026) demonstrated that explainable AI outputs supported faster analyst investigation by providing interpretable evidence rather than opaque scores. This principle is reflected in our per-dimension contribution scoring.

6.2 Limitations and Threats to Validity

First of all, it is simulation-based, and the parameters and statistics are tuned to match what was published as enterprise API utilization, and real-life deployments may have behavioral artifacts not covered by the model. The simulation must first be validated using actual OAuth telemetry before definitive recommendations can be made.

Second, the adversarial model does not account for advanced adversaries who perform abusive attacks after learning the behavioral profile of the victims through prior reconnaissance. Afzal and Huzaifa (2024) show how adaptive attackers can evade behavioral baselines through slow drift attacks. Future work can explore adversarially strong baseline estimation, including attention-based sequential modeling (Chandrappa et al., 2025).

To address the cold-start detection gap arising from requiring a minimum of 100 events per client identity to enable scoring (third feature), federated transfer learning (FTL) approaches (Chatterjee et al., 2022) can be leveraged to initialize baselines based on aggregate population statistics as per the multi-tenant data architecture framework (Chatterjee, 2023).

6.3 Privacy and Governance Considerations

The active OAuth telemetry collection processes user behavior data that is subject to privacy regulations like the EU GDPR, CA CCPA and other privacy laws. The framework has several mitigations in place for privacy compliance, including hashing of token identifiers, dropping raw IP addresses in favor of geolocation coordinates and extendable parameterized retention of telemetry. As Chatterjee (2023) states, governance documentation specific to that deployment of the monitoring pipeline should clearly define data stewardship roles, access control policies, and purpose limitation policies and procedures.

6.4 Integration with Existing Security Tooling

This scoring system is designed to work with existing security tools. Risk scores and feature contributions are exported as structured JSON events that can be consumed by tools like Splunk, Microsoft Sentinel, and IBM QRadar . Integration with identity governance controls can trigger automatic policy enforcement. If a token's risk score exceeds a revocation threshold, the mechanism calls the authorization server's token revocation endpoint, and also issues a step-up authentication challenge. Kakaraparthi (2022, cited in Kubam 2024) finds that as infrastructure asset management systems develop AI-based capabilities, the need for closed-loop security response automation will increase, and that high-velocity security incidents require design patterns of autonomous, real-time decision making with explainable outcomes.

7. CONCLUSION

This paper first provides two complementary artifacts to the problem of protecting OAuth tokens in cloud-native enterprise architectures. It introduces a 4D taxonomy of OAuth token abuse. This classification system effectively groups attacks by adoption vector, token type, privilege scope, and operational abuse pattern. This taxonomy enables

security practitioners to perform threat modelling and detection coverage assessments. The second contribution is an adaptive and behaviour-aware solution for continuous token lifecycle monitoring that aggregates OAuth telemetry and computes temporal and contextual features, adaptive EWMA baselines, and interpretable risk scores to ease automated response planning.

This yields a 94.2% detection rate in simulation with a 3.1% false positive rate. These rates are 21.8 and 12.5 percentage points higher than the authentication-only and rule-based token monitoring approaches, respectively. The geographic location of API requests is the most important detection feature, followed by the number of API calls and call scope. The EWMA baseline mechanism is resistant to behavioural drift from both legitimate users and slow-drift adversarial attacks, provided the parameters are properly chosen.

The contributions of this work are: a reusable taxonomy to classify rich authorization layer threats for analysis; a mathematical base-lining methodology for behavioural baseline, adapted from the well-studied field of statistical process control; and detection framework validated through simulation that is ready for integration with existing API gateway, SIEM, and identity governance tools. Future work will validate our framework against production OAuth telemetry, implement federated learning for cold start baseline initialization, and apply it to fine-grained OAuth scope graphs to detect privilege chaining attacks. With increasing footprints of enterprise SaaS applications and a move to zero trust architectures, the authorization layer is expected to become a greater focus for identity security programs.

REFERENCES

[1] Afzal, H., & Huzaifa, M. (2024). Securing AI systems against adversarial attacks: A framework for building robust and trustworthy machine learning models. Computer Fraud and Security. https://doi.org/10.52710/cfs.867

[2] Arella, S. G., & Malaraju, P. (2025). Secure session management in MERN stack applications using token rotation and refresh strategies. 2025 2nd International Conference on Recent Trends in Electrical, Electronics and Computing Technologies (ICRTEECT), Warangal, India, pp. 1–5. https://doi.org/10.1109/ICRTEECT67512.2025.11448621

[3] Chandrappa, S. K., Paheding, S., Reyes-Angulo, A. A., & Essa, A. (2025). Attention based spectral profile representation for hyperspectral image classification. NAECON 2025 – IEEE National Aerospace and Electronics Conference, Dayton, OH, USA, pp. 1–6. https://doi.org/10.1109/NAECON65708.2025.11235401

[4] Chatterjee, S. (2023). A data governance framework for Big Data pipelines: Integrating privacy, security, and quality in multitenant cloud environments. Technix International Journal for Engineering Research, 10(5), 746–758. https://doi.org/10.56975/tijer.v10i5.158181

[5] Shashank Lohani, The Learning Management System as a Product: A Computer Science Perspective on Higher Education Strategy, International Journal of Communication Networks and Information Security; vol 11 (2), 2019 page no 353-363

[6] Chatterjee, S. (2024). Understanding the risk of implementing AI for managing NERC BCSI data repository and security baseline to secure the BCSI data. Journal of Information Systems Engineering and Management, 9(4), 1–14. https://doi.org/10.52783/jisem.v9i4.19

[7] Giri, D. K., & Gangarapu, S. (2026). Explainable AI techniques for transparent algorithmic trading decisions. SoutheastCon 2026, Huntsville, AL, USA, pp. 01–08. https://doi.org/10.1109/SoutheastCon63549.2026.11475945

[8] Hardwick, F. S., Akram, R. N., & Markantonakis, K. (2019). Security vulnerabilities in OAuth 2.0 and their practical exploitation. In Proceedings of the 16th International Conference on Security and Cryptography (SECRYPT), pp. 303–314.

[9] Hardt, D. (Ed.). (2012). The OAuth 2.0 authorization framework (RFC 6749). Internet Engineering Task Force. https://datatracker.ietf.org/doc/html/rfc6749

[10] Kakaraparthi, G. C. (2022). A comparative study of LLMs for infrastructure-as-code generation and optimization. Computer Fraud & Security, 2022(4), 18–26. https://doi.org/10.52710/cfs.730

[11] Kubam, C. S. (2024). Agentic AI for autonomous, explainable, and real-time credit risk decision-making. Intelligent Systems and Applications in Engineering, 12(23). https://doi.org/10.5281/zenodo.18008872

[12] Lodderstedt, T., Bradley, J., Labunets, A., & Fett, D. (2020). OAuth 2.0 security best current practice (Internet-Draft). Internet Engineering Task Force. https://datatracker.ietf.org/doc/html/draft-ietf-oauth-security-topics

[13] Shashank Daram, "A Review of Retrieval-Augmented Generation in Natural Language Processing: Architectures, Challenges, and Future Research Directions" , International Journal on Recent and Innovation Trends in Computing and Communication; vol 8 (12), 2020 :145-155.

[14] Malaraju, S. K. (2026). An overview of cybersecurity risks and threats in cloud computing. In Lecture Notes in Networks and Systems (pp. 447–458). https://doi.org/10.1007/978-3-032-18211-1_40

[15] Rajendran, R. N., Anumula, S. K., Rai, D. K., & Agrawal, S. (2025). Zero trust security model implementation in microservices architectures using identity federation (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2511.04925

[16] Shernan, E., Carter, H., Tian, D., Traynor, P., & Butler, K. (2015). More guidelines than rules: CSRF vulnerabilities from noncompliant OAuth 2.0 implementations. In Proceedings of the 12th Conference on Detection of Intrusions and Malware & Vulnerability Assessment (DIMVA), pp. 239–260.

Author Note

All figures were generated using Python 3.11 with Matplotlib 3.8 and NumPy 1.26.