www.ijecs.in

International Journal Of Engineering And Computer Science
Volume 14 Issue 08 August 2025, Page No. 27619-27626
ISSN: 2319-7242 DOI: 10.18535/ijecs/v14i08.5207

Behavioral Authentication with Real-Time Anomaly Detection in Enterprise Logins

Sree Rajya Lakshmi Popury

Senior Engineer Consultant-Systems Engineering, Verizon Communications Inc.,
Dallas, Texas

Abstract

In this article, an approach to behavioral authentication with real-time anomaly detection during corporate system logins is examined. The objective of this study is to design and evaluate an architecture for continuous biometric user verification based on interaction dynamics, including cursor speed and trajectory curvature, inter-keystroke time intervals, and device micro-movements, while ensuring processing latency of below 100 ms under loads of up to tens of thousands of logins per second. It justifies because of the high percentage of incidents that happen out of credential compromise and low coverage of MFA, hence a continuous effort removal of friction for legitimate users. The novelty in this solution is introduced by the streaming architecture provided by Kafka and Flink-based applications, combined with telemetry normalization techniques using z-scores and robust scaling methods, together with an ensemble hybrid model consisting of a one-class autoencoder, relative attention, Isolation Forest, and a semi-supervised SSDLog scheme. Added value is brought by a dynamic threshold calibration mechanism that takes into account daily and weekly seasonality as well as federated learning and differential privacy noise mechanisms that satisfy privacy requirements. Findings present here will show that this proposed system ensures a median analysis latency of 26 ms, where the 99th percentile does not cross 51 ms, equal-error rate around about 1% at verification frequency of 0.7 s, reduces the number of interactive MFA challenges almost threefold for non-critical accounts and ensures adaptation to user behavior drift through continuous self-learning and feedback from SOC analysts. This article will be helpful to information security specialists, authentication system developers, and IT architects of large corporate infrastructures.

Keywords: behavioral authentication, stream processing, Kafka, Flink, autoencoder, anomaly, MFA

Introduction

Password remains the most frequent entry point for adversaries: in 2023, the use of stolen credentials was the initial action in 24% of confirmed breaches, and over the past ten years, such data have appeared in nearly one-third of incidents (Verizon, 2024). Corporate password hardening policies reduce convenience but do not eliminate the issue of password reuse. A recent FIDO Alliance study revealed that 51% of users reuse the same passwords across various resources, and 59% of employees rely solely on a password for workplace access (Zaky, 2024).

Multi-factor authentication significantly mitigates the risk of compromise; yet, its adoption remains insufficient. Microsoft's telemetry analysis revealed that more than 99.9% of compromised corporate accounts did not have MFA enabled (Microsoft, 2024a). Such statistics indicate a structural vulnerability in the classic login, password, and optional MFA scheme, as protection is activated only upon an explicit factor request, which is often postponed or canceled by the user.

Behavioral authentication offers an alternative approach: the system builds a unique user profile from typing dynamics, cursor trajectory, device micro-movements, and other hard-to-forge markers collected by sensors in the background. Contemporary research defines this technology as continuous biometric verification based on embedded sensors that requires no additional hardware (Finnegan et al., 2024).

Verification occurs unobtrusively and continuously, so the probability that an adversary will replicate the entire behavioral template is significantly lower than guessing a password or intercepting a one-time code.

Thus, the high share of password-based breaches and low MFA saturation make behavioral authentication a timely supplementary security layer capable of reducing incident numbers without increasing friction for legitimate employees.

Materials and Methodology

The study is based on the analysis of 17 sources, including academic articles, technical reports, industry white papers, and case studies. The theoretical foundation comprises works on continuous biometric verification using typing dynamics, cursor trajectory and sensor data (Finnegan et al., 2024), NIST SP 800-63B recommendations on risk-based authentication (Grassi et al., 2017), as well as Microsoft reports on MFA usage statistics in corporate environments (Microsoft, 2024a; Microsoft, 2024b). To understand the scalability and reliability of streaming architectures, studies examining the performance of Apache Kafka and Flink under high loads were conducted (Cumbane & Gidófalvi, 2019; Confluent, 2025).

The methodological approach combined several complementary stages. A comparative analysis of telemetry collection and transmission technologies included an evaluation of Kafka's at-least-once semantics versus Flink's exactly-once semantics and Spark Streaming, with measurements of median latency and 99th percentile under loads of up to tens of thousands of logins per second (Cumbane & Gidófalvi, 2019; Confluent, 2025). A systematic review of data preprocessing techniques covered hold time and flight time computation for keyboard events, mouse trajectory curvature, and spectral coefficients of sensor signals, followed by normalization via z-score and robust scaling (Shadman, 2025).

Results and Discussion

A corporate agent, embedded in a web client or mobile application, begins collecting telemetry even before the user clicks the Sign In button. The stream captures inter-keystroke intervals, cursor trajectories, device angular accelerations, and network metadata. Raw events are transmitted to a Kafka queue, which guarantees ordering and provides horizontal scalability, then processed in a Flink or Spark Streaming engine. Field tests show that when handling tens of thousands of logins per second, Flink maintains a median latency of 26 ms and does not exceed 51 ms at the 99th percentile, satisfying interactive UX requirements (Cumbane & Gidófalvi, 2019).

In the next stage, the stream is transformed into feature vectors. For keyboard data, for example, hold time, flight time, and the ratio of their total durations are computed. Similarly, for mouse and sensor data, speeds, curvature, and spectral coefficients are derived. Such a feature set has demonstrated robustness to contextual variability and ease of subsequent normalization, since most metrics reduce to time distributions measured in milliseconds (Shadman, 2025). Z-score or robust scaling is applied to the data before it is fed into the model.

A behavioral profile is constructed on an individual basis: the system trains one or more unsupervised models on the user's legitimate sessions. In practice, a one-class autoencoder with relative attention performs well by reconstructing a normal representation and using reconstruction error as a risk indicator. This method achieves an approximately 1% equal-error rate at a 0.7-second verification frequency, does not require negative examples, and maintains low parameter dimensionality (Hu et al., 2022). For service accounts with scant data, the profile becomes even more enriched with an aggregated template of the team or department. In use, inference runs over every sliding window of telemetry, and the engine gives out a numerical score. Lightweight Hoeffding trees or adaptive thresholds enable evaluation latency of below 100 ms, even under concept drift, while computational costs remain acceptable for edge nodes (Confluent, 2025). All scores and key metrics (including outlier count and event density) are stored for offline analytics and retraining purposes.

A policy orchestrator makes the final decision. It correlates the risk score with session context, employee role and application type, following NIST SP 800-63B recommendations on risk-based authentication: for instance, a request from a previously known IP range with a low score is passed without additional factors, whereas an atypical geolocation and a high score automatically escalate to WebAuthn or blocking (Grassi et al., 2017). Such cascade intervention minimizes false alarms, preserves workflow continuity, and simultaneously ensures rapid defense against session hijacking.

Since the authentication stream begins before the user sees the login form, the architecture integrates passive telemetry collection into the client session itself. A lightweight agent in the browser and mobile container subscribes to keyboard, mouse, touch sensor, and network stack events, buffering them in the event loop, which eliminates the need for additional requests.

Subsequently, data are published to a distributed message bus. In the reference implementation, Apache Kafka is used with segmentation by tenant ID and geographic domain. This configuration, validated by LinkedIn's production deployment, reliably handles peaks exceeding 13 million messages per second across over a thousand brokers with a daily volume of roughly 800 billion events, thereby setting the upper bound for the entire pipeline's throughput (Palino, 2015). Thus, even large corporate domains can maintain all login events in firehose mode without risk of failure.

At the stream-processing layer, a Flink cluster with Exactly-Once semantics enabled ensures that the risk assessment is returned before rendering the final page or triggering an additional factor, thereby keeping the total interaction duration within the user's expected pause between clicking and transitioning.

The risk-evaluation block constructs a feature vector via a sliding window, normalizes it, and submits it to an ensemble of models. A seasonal autoencoder base tracks individual patterns, overlaid by a gradient booster trained on simulated attacks. The threshold decision is tied to access policies; exceeding the acceptable risk triggers a step-up, for example, WebAuthn plus PIN. This approach aligns with NIST SP 800-63B's allowance for adaptive authentication based on spatio-temporal anomalies and session metadata (Grassi et al., 2017).

Each decision, together with the complete feature set and telemetry, is recorded in immutable storage. A nightly offline pipeline retrains models while accounting for daily and weekly shifts in employee behavior. A control sample from the open KBOC dataset is used to monitor accuracy degradation. Even a simple Manhattan detector yields an EER of 5.32% on that dataset, which serves as a working benchmark for internal backend tests (Monaco, 2016). Delayed feedback is returned to agents as updated filtering rules, thereby closing the system's self-learning loop.

In constructing the anomaly-detection layer, the system relies on three complementary families of algorithms, each addressing a specific task. The first group pertains to classical unsupervised analysis, necessary during initial deployment stages when attack labels are scarce. On corporate keyboard datasets, Isolation Forest after minimal tuning achieves an equal-error rate of 7.81 % while maintaining an area under the ROC curve of 0.97, thus providing an acceptable trade-off between false blocks and missed intrusions without requiring labeled data, as shown in Figure 1 (Ismail et al., 2024).

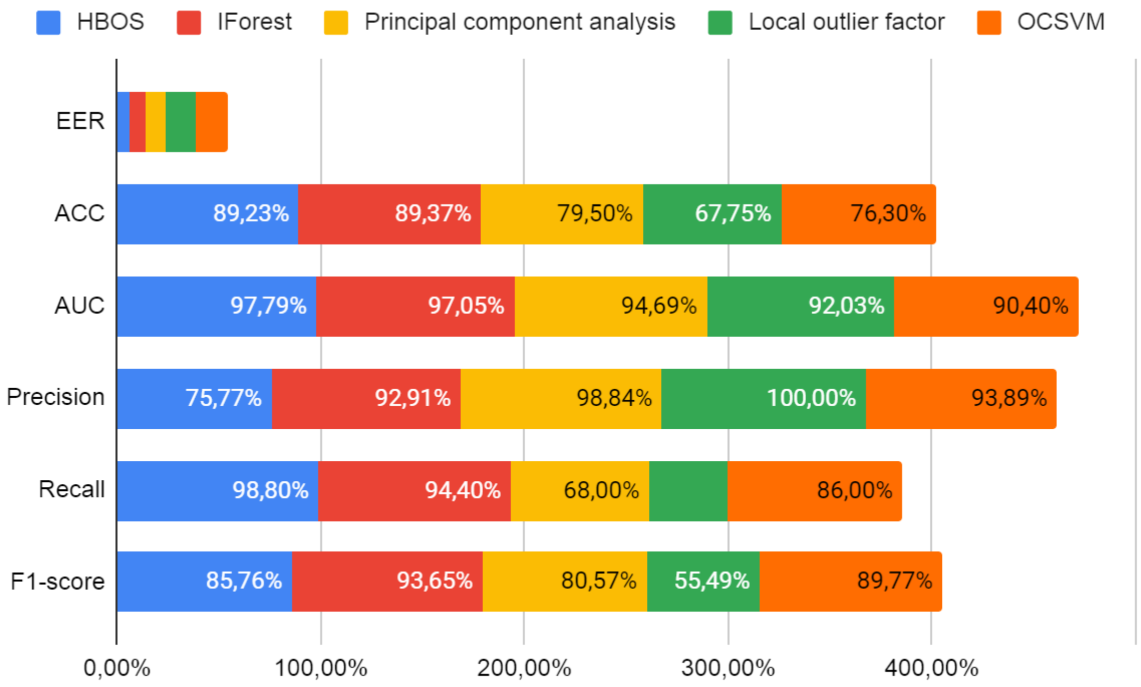

Fig. 1. Comparison of five tuned outlier detection models in keystroke biometric authentication using the CMU dataset (Ismail et al., 2024)

Deeper models, such as a relative-attention one-class autoencoder trained only on clean sessions, reduce the EER to 1.05% at a verification frequency of 0.7 s, thereby allowing for the practical elimination of additional multi-factor authentication in normally operating pipelines (Hu et al., 2022).

As labeled incidents gradually accumulate in the logs, a semi-supervised scheme is engaged. Its typical representative is SSDLog; the model uses only 30% of the labeled log lines, the remainder are automatically pseudo-labeled, and training proceeds in a teacher-student configuration. On the HDFS dataset, this strategy increased the F1-score to 98.7% and raised detection recall to 99.4%, thereby nearly matching a fully supervised solution while incurring one-third the expert effort (Lu et al., 2023). Such results indicate that, in a corporate setting where the SOC records every confirmed intrusion, within a few weeks, one can transition from general profiles to finely tuned detectors that are specifically sensitive to attacks on that organization.

The third avenue combines different models into hybrid ensembles, enhancing robustness against attempts to emulate legitimate user behavior. A recent study of a privacy-oriented ensemble combining KNN, SVM, XGBoost, and a lightweight neural network reports a final accuracy of 94.3 % and an F1-score of 93.5 %, outperforming all individual base models; moreover, the architecture incorporates a differential-privacy noise mechanism and federated learning to comply with internal personal-data handling policies, as shown in Figure 2 (Liu & Zhao, 2023).

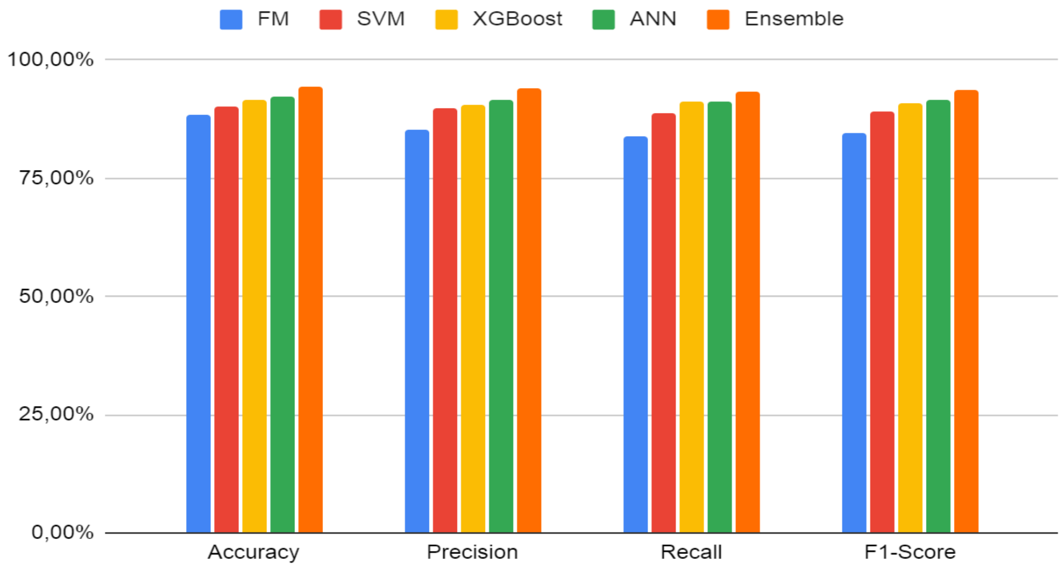

Fig. 2. Performance Comparison of Different Models (Liu & Zhao, 2023)

In production environments, such a cascaded strategy is often implemented in two tiers: fast filtering of the stream by Isolation Forest or autoencoder at the edge, and a heavier ensemble in the core, to which only suspicious sessions are forwarded.

The combined use of these three layers strikes a balance between response speed and accuracy. Unsupervised models guarantee zero-day coverage, semi-supervised models gradually reduce false alarms, and ensembles provide resilience against evasion through error diversification. All algorithms output a unified risk score, which the policy orchestrator then leverages to link the machine decision with business context and compliance requirements.

Threshold tuning begins with the system refraining from setting a single risk value; instead, it dynamically adjusts limits based on recent observations. In practice, the sliding window in corporate networks spans 300–500 score values, after which the boundaries are recalculated using Shewhart rules: a point is deemed anomalous when it falls outside the three-sigma corridor relative to the local mean.

Employee behavior exhibits strong daily and weekly seasonality, so the model is augmented with calendar features: hour of day and day-type are encoded as cyclical vectors; for rare holidays, a regressor trained on a multiyear log of 1,910 days of real cloud-storage operation is applied, where deviations during nonworking hours exceeded the baseline by fourfold (Landauer et al., 2022). This adjustment reduces evening and weekend false-alarm spikes without diminishing attack sensitivity, since anomalies are assessed relative to the local, not global, distribution.

Next, the orchestrator applies risk stratification. An account with Global Admin privileges or access to financial systems is automatically placed in the high class, where even a medium-risk score triggers additional verification. For ordinary users, the same score may remain within monitor without intervention. Microsoft Entra documentation notes that raising the threshold from Medium to High reduces interactive MFA prompts by nearly threefold while preserving protection for key accounts through stricter policies (Microsoft, 2024b).

The security-usability balance is achieved through a cascaded factor escalation. At low or medium risk, the session concludes without extra steps; at high risk, the system requests visible MFA. According to Microsoft research, proper multi-factor enforcement blocks over 99.2% of credential-compromise attempts, so precise escalation timing both reduces user burden and maintains a high barrier for attackers (Microsoft, 2025).

To build SOC analysts' trust in the automated decision, the engine returns not only the numeric risk but also a brief rationale, which includes the features that contributed most, the control limits that were exceeded, and the influence of the seasonal corrector. Thus, dynamic thresholds, calendar-effect adjustment, and contextual stratification form a closed-loop system in which security is strengthened while user friction is systematically reduced, as illustrated in Figure 3.

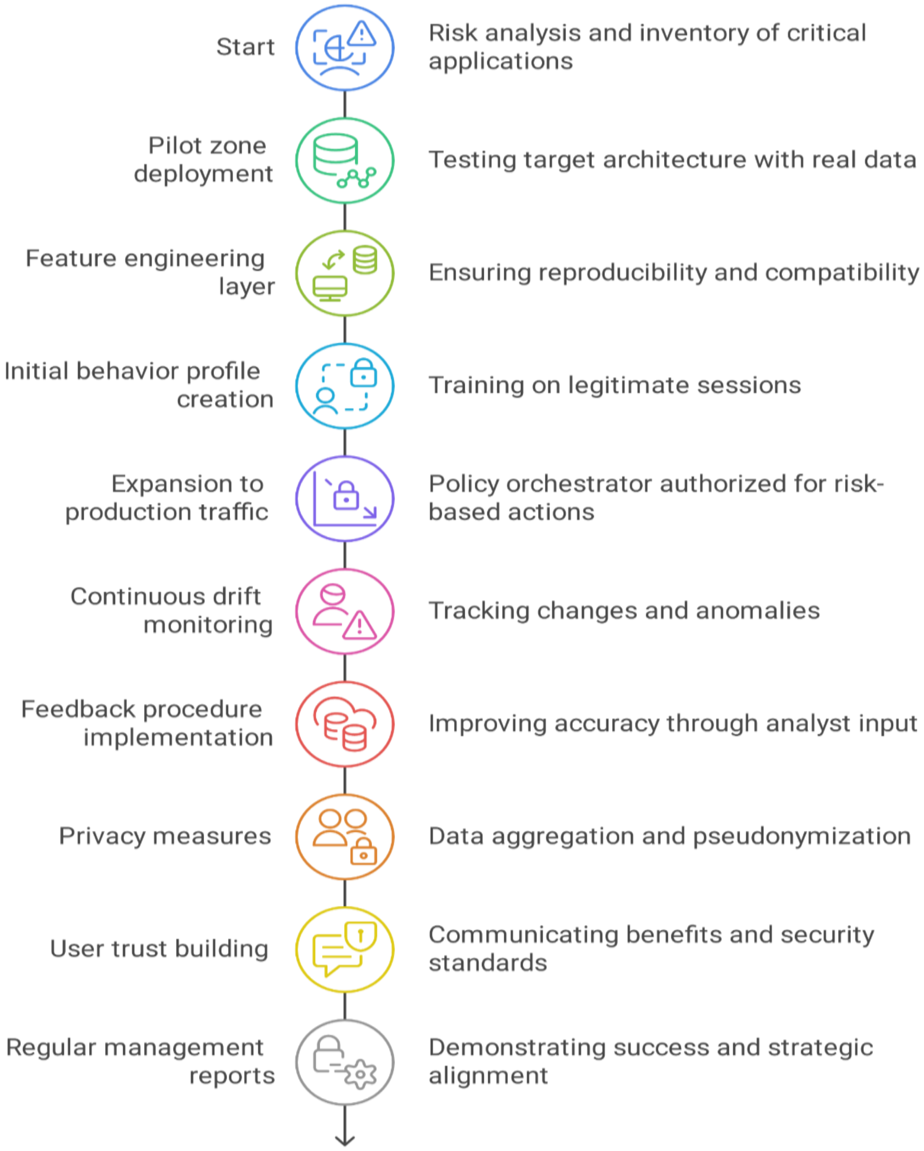
Start
Risk analysis and inventory of critical applications
Pilot zone deployment
Testing target architecture with real data
Feature engineering layer
Ensuring reproducibility and compatibility
Initial behavior profile creation
Training on legitimate sessions
Expansion to production traffic
Policy orchestrator authorized for risk-based actions
Continuous drift monitoring
Tracking changes and anomalies
Feedback procedure implementation
Improving accuracy through analyst input
Privacy measures
Data aggregation and pseudonymization
User trust building
Communicating benefits and security standards
Regular management reports
Demonstrating success and strategic alignment

Fig. 3. A Systematic Framework for the Deployment of Behavioral Authentication Mechanisms (compiled by author)

The implementation of behavioral authentication is most appropriately initiated with a risk analysis and an inventory of mission-critical applications, as the depth of telemetry and the level of control depend directly on them. At this stage, requirements are defined for latency, event throughput, and regulatory context, and it is clarified which user groups require the highest level of protection. Next, the target architecture is deployed in a pilot zone, where real streams of keyboard, mouse, and network metrics are collected, with the agent embedded into existing interfaces without altering login logic; this enables the performance of the message broker and the stream-processing engine to be tested with minimal impact on employee experience.

Once data ingestion is stabilized, the team isolates feature engineering as a distinct layer, ensuring the reproducibility of feature extraction and compatibility across multiple models. The transformation mechanism must support versioning so that historical sessions remain comparable to new data when sensors are replaced or added. Primary behavioral profiles are then created from the resulting feature vectors, which are trained exclusively on known legitimate sessions. At this point, it is crucial to establish a manual validation procedure to exclude latent attacks from the training set. After threshold calibration, the pilot zone is extended to production traffic, and the policy orchestrator is granted authority to request multi-factor authentication or to block access based on evaluated risk.

The next task is continuous drift monitoring. The system must automatically track changes in statistical distributions and alert when the volume of anomalies exceeds confidence intervals, indicating a need for retraining or threshold adjustment. In parallel, a feedback procedure from the security operations center is implemented: analysts label sessions as confirmed incidents or false alarms, and an active-learning mechanism incorporates these labels into the repository. This process progressively improves model accuracy and reduces the frequency of escalations.

Special attention is paid to privacy. Before storage, data are aggregated and pseudo-anonymized, and privacy-preserving techniques can be employed during model training to mitigate the influence of any individual session. Corporate policies codify the principle of data minimization: in most scenarios, derived features suffice, so raw cursor coordinates or full network packets are purged immediately after feature extraction. Consequently, even in the event of a data breach, the telemetry does not disclose sensitive employee information.

Finally, success depends on user trust. Communication must clarify that the additional security layer operates unobtrusively and reduces the number of explicit challenges, while all processes remain strictly governed by internal security standards. Regular management reports should demonstrate reductions in attacks and time saved on manual verification, thereby embedding the initiative within corporate strategy and securing resources for further system development.

Conclusion

The present study demonstrates that behavioral authentication based on continuous collection and analysis of user-interaction telemetry can substantially enhance the security of corporate systems without appreciably increasing friction for legitimate users. Integration of a lightweight agent into a web client or mobile application enables the capture of numerous features—from inter-keystroke intervals to mouse-trajectory curvature and spectral coefficients of sensor data—already during the loading of the authentication page. Use Flink or Spark Streaming as an implementation of Stream processing at a median delay of 26 ms, where the 99th percentile does not take more than 51 ms, well within the boundaries of an interactive UX and real-time analysis at loads that can reach tens of thousands of logins per second.

The modeling layer is based on a one-class autoencoder with relative attention, achieving about 1% error for verification that takes less than a second. This approach is presented without negative examples, thereby maintaining the compactness of the parameter dimensionality. The ensemble approach, at the core of the system, balances the reaction speed that can be achieved at the edge level for an incoming stream with the more complex hybrid models, while maintaining detection accuracy. Deploying Isolation Forests at initial deployment together with SSDLog semi-supervised schemas using pseudo-labeling permits both coverage of zero-day attacks and adaptation to corporate-specific needs. These are adjusted dynamically to threshold-tuning mechanisms that account for seasonal and calendar effects, resulting in minimal false-positive rates.

A policy orchestrator, guided by NIST SP 800-63B standards, correlates the risk score with session context, user role, and application type, allowing escalation of challenges only for truly anomalous events:

requests from familiar IP ranges proceed transparently, whereas atypical geolocations or elevated scores trigger step-up authentication or blockage. This cascaded strategy reduces interactive multi-factor authentication prompts nearly threefold for non-critical accounts while maintaining robust protection for key roles.

Continuous system self-retraining, implemented via nightly offline pipelines and feedback from SOC analysts, ensures adaptation to evolving user behavior and maintains model relevance. Meanwhile, pseudo-anonymization and telemetry minimization procedures satisfy corporate and regulatory requirements for personal data protection. Through a comprehensive approach—including multilayered detection, dynamic thresholds, and transparent reporting to SOC teams—the deployment of behavioral authentication becomes an effective supplementary defense line, capable of significantly reducing password-compromise incidents without degrading the user experience.

References

1. Confluent. (2025). *Delivery Guarantees and Latency in Confluent Cloud for Apache Flink*. Confluent. https://docs.confluent.io/cloud/current/flink/concepts/delivery-guarantees.html
2. Cumbane, S. P., & Gidófalvi, G. (2019). Review of Big Data and Processing Frameworks for Disaster Response Applications. *ISPRS International Journal of Geo-Information*, *8*(9), 387. https://doi.org/10.3390/ijgi8090387
3. Finnegan, O. L., White, J. W., Armstrong, B., Adams, E. L., Burkart, S., Beets, M. W., Willis, E. A., Parker, H., Bastyr, M., Zhu, X., Zhong, Z., & Weaver, R. G. (2024). The utility of behavioral biometrics in user authentication and demographic characteristic detection: a scoping review. *Systematic Reviews*, *13*(1). https://doi.org/10.1186/s13643-024-02451-1
4. Grassi, P. A., Fenton, J. L., Newton, E. M., Perlner, R. A., Regenscheid, A. R., Burr, W. E., Richer, J. P., Lefkovitz, N. B., Danker, J. M., Choong, Y.-Y., Greene, K. K., & Theofanos, M. F. (2017). Digital Identity Guidelines: Authentication and Lifecycle Management. *NIST Special Publication 800-63B*. https://doi.org/10.6028/nist.sp.800-63b
5. Hu, M., Zhang, K., You, R., & Tu, B. (2022). Relative Attention-based One-Class Adversarial Autoencoder for Continuous Authentication of Smartphone Users. *Arxiv*. https://doi.org/10.48550/arxiv.2210.16819
6. Ismail, M. G., Salem, M. A.-M., Abd, M. A., & Abbas, S. (2024). Outlier detection for keystroke biometric user authentication. *PeerJ Computer Science*, *10*, e2086–e2086. https://doi.org/10.7717/peerj-cs.2086
7. Landauer, M., Skopik, F., Höld, G., & Wurzenberger, M. (2022, December 1). *A User and Entity Behavior Analytics Log Data Set for Anomaly Detection in Cloud Computing*. IEEE Xplore. https://doi.org/10.1109/BigData55660.2022.10020672
8. Liu, S., & Zhao, Z. (2023). *Privacy-Preserving Hybrid Ensemble Model for Network Anomaly Detection: Balancing Security and Data Protection*. Arxiv. https://ar5iv.labs.arxiv.org/html/2502.09001
9. Lu, S., Han, N., Wang, M., Wei, X., Lin, Z., & Wang, D. (2023). SSDLog: a semi-supervised dual branch model for log anomaly detection. *World Wide Web*, *26*(5), 3137–3153. https://doi.org/10.1007/s11280-023-01174-y
10. Microsoft. (2024a, May 13). *Security at your organization - Multifactor authentication (MFA) statistics*. Microsoft. https://learn.microsoft.com/en-us/partner-center/security/security-at-your-organization
11. Microsoft. (2024b, June 17). *Risk policies - Microsoft Entra ID Protection*. Microsoft. https://learn.microsoft.com/en-us/entra/id-protection/howto-identity-protection-configure-risk-policies
12. Microsoft. (2025, February 28). *Plan for mandatory Microsoft Entra multifactor authentication (MFA) - Microsoft Entra ID*. Microsoft. https://learn.microsoft.com/en-us/entra/identity/authentication/concept-mandatory-multifactor-authentication?tabs=dotnet
13. Monaco, J. V. (2016). Robust Keystroke Biometric Anomaly Detection. *Arxiv*. https://doi.org/10.48550/arxiv.1606.09075
14. Palino, T. (2015). *Running Kafka At Scale*. LinkedIn. https://engineering.linkedin.com/kafka/running-kafka-scale

15. Shadman, R. (2025). *Keystroke Dynamics: Concepts, Techniques, and Applications*. Arxiv. https://arxiv.org/html/2303.04605v3
16. Verizon. (2024). *2024 Data Breach Investigations Report*. Verizon. https://www.verizon.com/business/resources/reports/2024-dbir-data-breach-investigations-report.pdf
17. Zaky, K. (2024, August 29). *White Paper: Replacing Password-Only Authentication with Passkeys in the Enterprise*. FIDO Alliance. https://fidoalliance.org/white-paper-replacing-password-only-authentication-with-passkeys-in-the-enterprise/