Behavioural Profiling-Based Authentication Anomaly Detection using a Data-Driven Framework

K. Anuranjani
Department of Computer Science and Engineering
Bharath Institute of Higher Education and Research Tambaram, Chennai
anuranjani.cse@bharathuniv.ac.in

Magesh T N
Department of Computer Science and Engineering
Bharath Institute of Higher Education and Research
Chennai, India
magesh4214@gmail.com

Dinesh Babu V
Department of Computer Science and Engineering
Bharath Institute of Higher Education and Research
Chennai, India
venkatasandinesh@gmail.com

Lydian Joshwa J
Department of Computer Science and Engineering
Bharath Institute of Higher Education and Research
Chennai, India
lydianjoshwa@gmail.com

***Abstract*—In our digital world, keeping accounts safe is a constant battle. While we rely on passwords and two-factor codes, these methods only check who you are at the front door. They don't notice if someone else sneaks in using stolen keys. This project introduces a smarter way to protect users: a data-driven framework that gets to know your unique digital "personality." By analyzing habits like when you usually log in, where you're located, and what device you use, the system builds a secure behavioral profile. Using machine learning algorithms like Random Forest and Neural Networks, it learns to spot the difference between your normal routine and suspicious activity in real-time. If something feels off—like a login from a new country at 3 AM—the system immediately steps in, triggering alerts or extra security checks to stop attackers in their tracks. Our tests show this approach is highly accurate, catching anomalies without making life difficult for legitimate users. It's a scalable, adaptive solution designed to stay ahead of evolving cyber threats. By shifting from static passwords to intelligent, continuous monitoring, this research proves that we can create a digital environment that is both more secure and more intuitive for everyone.**

Keywords—Keywords—Anomaly Detection, Machine Learning, Authentication

I. INTRODUCTION

Modern digital systems increasingly rely on secure authentication mechanisms to protect user accounts and sensitive data. With the rapid growth of online services, cloud platforms, and digital transactions, authentication systems have become prime targets for cyberattacks such as credential theft, phishing, brute-force attacks, and account takeover incidents. Traditional authentication methods, including username-password combinations, One-Time Passwords (OTP), and CAPTCHA-based verification, primarily validate user identity only at the initial login stage. However, these approaches lack continuous monitoring capabilities, making them vulnerable when attackers gain access to valid credentials.

To address these limitations, there is a growing need for intelligent and adaptive authentication systems that extend beyond static verification methods. Behavioural profiling has emerged as a promising approach, where systems analyze user-specific patterns such as login time, device type, IP address, geographic location, and usage frequency to establish a unique behavioural signature. By continuously monitoring these patterns, deviations from normal behaviour can be identified, enabling early detection of suspicious or malicious activities. In this context, this paper proposes a behavioural profiling-based authentication framework for anomaly detection using machine learning techniques. The system leverages historical authentication data to model normal user behaviour and evaluates real-time login attempts against these learned patterns. Machine learning algorithms such as Logistic Regression, Random Forest, Support Vector Machine (SVM), and Neural Networks are employed to classify login activities as normal or anomalous. This data-driven approach enables accurate detection of unauthorized access, even when valid credentials are used. Furthermore, the proposed system supports real-time anomaly detection and incorporates a security response mechanism that triggers alerts, additional authentication, or access restrictions upon detecting suspicious behaviour. By integrating behavioural analytics with machine learning, the framework enhances authentication security while maintaining a balance between protection and user experience. This approach provides a scalable and efficient solution for modern digital environments, including enterprise systems, financial platforms, and cloud-based applications.

II. EASE OF USE

A. System Usability and Design

The proposed behavioural profiling-based authentication system is designed with a strong emphasis on usability and seamless integration into existing digital platforms. Unlike

traditional authentication mechanisms that require frequent user interaction through repeated password entries, OTP verification, or CAPTCHA challenges, the proposed system operates primarily in the background. It continuously monitors user behaviour based on parameters such as login time, device type, IP address, geographic location, and login frequency, without requiring additional effort from the user. This non-intrusive approach enhances the overall user experience by reducing unnecessary interruptions while maintaining a high level of security. Legitimate users can access the system smoothly under normal behavioural conditions, whereas only suspicious activities trigger additional verification steps. As a result, the system effectively balances security requirements with user convenience.

B. Integration and Adaptability

The framework is designed to be easily integrated into modern web applications, enterprise systems, and cloud-based platforms. Since the system relies on authentication logs that are already generated in most applications, it does not require major modifications to existing infrastructure. The modular architecture, consisting of data collection, preprocessing, model training, anomaly detection, and security response components, allows flexible deployment and scalability. Furthermore, the system is adaptable to dynamic user behaviour. Machine learning models are periodically updated using new authentication data, enabling the system to learn evolving user patterns and maintain detection accuracy over time. This adaptability ensures that the system remains effective against emerging cyber threats and changing usage patterns.

C. Real-Time Operation and Efficiency

The proposed system supports real-time anomaly detection, ensuring that each login attempt is evaluated instantly. This enables immediate identification of suspicious activities and timely activation of security responses such as alerts, additional authentication, or access restriction. To maintain efficiency, the system utilizes optimized machine learning models that provide accurate predictions with minimal computational overhead. This makes it suitable for deployment in large-scale environments with high user activity, where quick decision-making and low latency are critical.

III. LITERATURE SURVEY

The increasing reliance on digital platforms has intensified the need for robust authentication mechanisms capable of defending against sophisticated cyber threats such as credential theft, phishing, and account takeover. Traditional authentication systems, primarily based on passwords and one-time verification techniques, have proven insufficient due to their static nature and lack of continuous monitoring. This limitation has led to the exploration of anomaly detection and behaviour-based authentication methods as more effective security solutions.

Early research in anomaly detection focused on statistical techniques and unsupervised learning methods to identify deviations from normal patterns without requiring labeled attack data. These approaches demonstrated effectiveness in detecting rare and abnormal events, making them suitable for security applications where malicious activities are infrequent but critical. With advancements in machine learning, supervised algorithms such as Logistic Regression, Support Vector Machines (SVM), and Random Forest have been widely adopted for classifying user behaviour. These models offer improved accuracy by learning complex patterns from historical data and adapting to evolving threats.

Behavioral biometrics has further enhanced authentication systems by introducing continuous monitoring based on user-specific patterns such as login time, device usage, IP address, and geographic location. Unlike traditional methods that verify identity only at login, behavioural profiling enables systems to validate users throughout a session, improving detection of unauthorized access even when valid credentials are used. However, challenges such as variability in user behaviour, high false positive rates, and privacy concerns remain significant.

Recent studies have also explored deep learning techniques, including neural networks, to capture complex and non-linear behavioural patterns. While these approaches provide higher detection capability, they often require large datasets and increased computational resources. Additionally, hybrid models that combine multiple machine learning techniques have shown promising results in improving detection accuracy and robustness by leveraging the strengths of different algorithms.

Despite these advancements, several research gaps persist, including limited implementation of continuous authentication, insufficient integration of multiple behavioural features, challenges in handling imbalanced datasets, and lack of real-time scalability. These limitations highlight the need for a comprehensive, data-driven authentication framework. The proposed system addresses these gaps by integrating behavioural profiling with multiple machine learning models to achieve accurate, adaptive, and real-time anomaly detection.

IV. METHODOLOGY

The proposed system introduces a behavioural profiling-based authentication framework that utilizes machine learning techniques to detect anomalous login activities in real time. The methodology follows a structured pipeline consisting of data collection, preprocessing, feature engineering, model training, anomaly detection, and security response.

A. System Overview

The system continuously monitors user authentication behaviour by analyzing parameters such as login time, IP address, device type, geographic location, browser type, and login frequency. These attributes are used to construct a behavioural profile for each user, representing normal activity patterns. Any deviation from this profile is treated as a potential anomaly.

B. Data Collection and Preprocessing

Authentication logs are collected from the system, containing relevant user activity data. The collected data

undergoes preprocessing to ensure quality and consistency. This includes removal of duplicate entries, handling missing values, and transforming raw data into a structured format suitable for analysis.

Categorical features such as device type and browser are encoded using appropriate techniques, while numerical features such as login time and frequency are normalized. This step ensures that the dataset is clean and ready for effective model training.

C. Feature Engineering

Feature engineering is performed to extract meaningful information from the authentication data. Temporal features such as login hour and login intervals are derived from timestamps, while location-based and device-based attributes are encoded to capture user access patterns.

Additional features such as login frequency and IP variation are also considered to enhance anomaly detection capability. These features collectively form the input vector used by machine learning models.

D. Model Selection and Training

Multiple machine learning algorithms are employed to model user behaviour, including Logistic Regression, Random Forest, Support Vector Machine (SVM), and Neural Networks. Each model is trained using historical authentication data to learn normal behavioural patterns.

The dataset is divided into training and testing sets to evaluate model performance. Techniques such as hyperparameter tuning and cross-validation are applied to improve accuracy and generalization. Additionally, data imbalance is addressed using appropriate resampling techniques to ensure effective anomaly detection.

E. Anomaly Detection Mechanism

During real-time operation, each login attempt is processed and converted into a feature vector consistent with the training data. The trained models analyze this input and compare it with the learned behavioural profile.

A decision threshold is used to classify the login attempt as normal or anomalous. In cases where multiple models are used, a combined decision approach such as majority voting is applied to improve detection accuracy.

F. Security Response Mechanism

Upon detection of anomalous behaviour, the system triggers appropriate security responses based on the severity of the deviation. These responses may include generating alerts, requesting additional authentication (e.g., OTP), or temporarily restricting access.

The system also maintains logs of detected anomalies for further analysis and continuous improvement. This ensures enhanced security while minimizing inconvenience to legitimate users.

V. Results and Discussions

A. Dataset Visualization

The dataset used in this study consists of user authentication logs containing behavioural attributes such as login time, IP address, device type, geographic location, and login frequency. Understanding the distribution and characteristics of this data is essential for building an effective anomaly detection model.

The distribution of login behaviour is illustrated in Fig. 1, which presents a density plot of the dataset. The figure highlights variations in user activity patterns and helps identify regions where anomalous behaviour deviates from normal usage. This initial analysis provides insights into the underlying data structure and supports the effectiveness of machine learning-based classification.

The dataset is further visualized in Fig. 2, where patterns of normal and anomalous login activities are represented graphically. The visualization demonstrates clear separability between behavioural clusters, indicating that the selected features are effective in distinguishing legitimate users from potential threats. Such visual insights validate the suitability of the dataset for anomaly detection tasks.

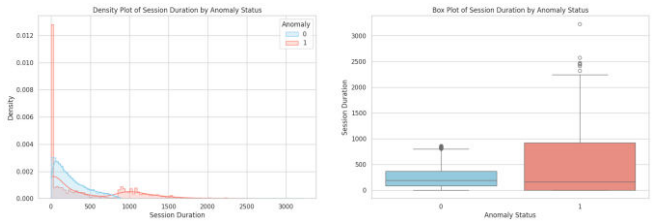

Fig. 1 Density plot of dataset distribution

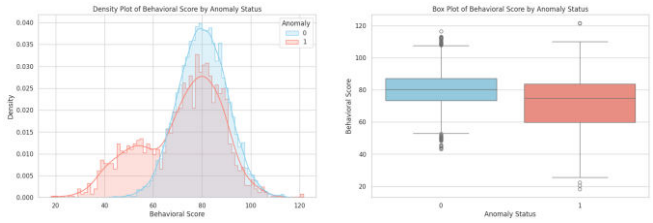

Fig. 2 Visualization of login behaviour patterns

B. Model Performance Comparison

To evaluate the effectiveness of the proposed anomaly detection system, multiple machine learning models were implemented and analyzed, including Logistic Regression, Random Forest, Support Vector Machine (SVM), and Neural Networks. Each model was trained using historical authentication data and tested on unseen data to assess its ability to accurately classify login attempts as normal or anomalous.

Among the evaluated models, the Random Forest classifier demonstrated superior performance in terms of accuracy and robustness. Its ensemble learning approach, which combines multiple decision trees, enables it to effectively capture complex behavioural patterns and reduce overfitting. This makes it particularly suitable for handling diverse authentication features such as login time, device type, IP address, and geographic location.

The classification results obtained using the Random Forest model are illustrated in Fig. 3. The figure shows a clear distinction between normal and anomalous login activities, highlighting the model's ability to accurately detect deviations in user behaviour. This improved detection capability contributes to enhanced security by identifying suspicious activities in real time while maintaining low false positive rates.

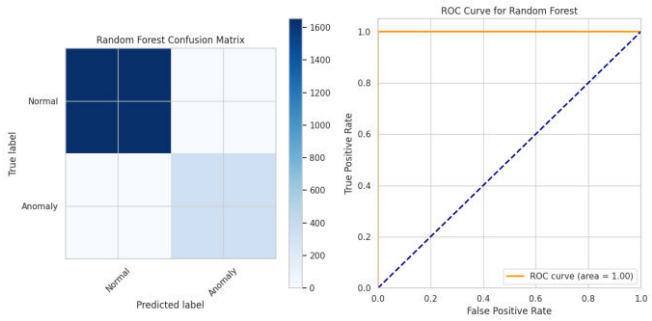

Fig. 3 Random Forest classification output

C. Output Visualization

In addition to traditional machine learning models, a Neural Network-based approach was implemented to capture complex and non-linear behavioural patterns. Neural Networks are particularly effective in learning intricate relationships within data, making them suitable for advanced anomaly detection scenarios.

The output of the Neural Network model is presented in Fig. 4. The figure demonstrates the model's capability to classify login attempts with a high degree of accuracy, highlighting its effectiveness in identifying subtle anomalies that may not be easily detected by simpler models.

Although the Neural Network provides strong performance, it requires higher computational resources compared to other models. Therefore, it is best suited for environments where advanced detection capability is prioritized alongside system scalability.

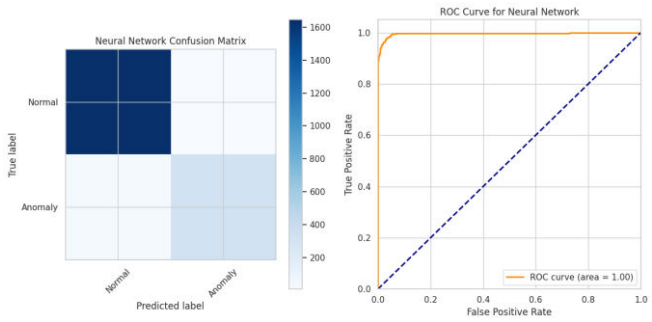

Fig. 4 Neural Network classification output

D. Discussion

The experimental results indicate that the proposed behavioural profiling-based authentication system is effective in detecting anomalous login activities. The combination of multiple machine learning models enables accurate classification while maintaining a balance between security and user experience.

Random Forest emerged as the most reliable model due to its robustness and efficiency, while Neural Networks provided deeper insights into complex behavioural patterns. The visualizations and performance analysis collectively demonstrate that integrating behavioural analytics with machine learning significantly enhances authentication security.

Overall, the system achieves high detection accuracy with reduced false positives, making it suitable for deployment in real-world applications such as enterprise systems, financial platforms, and cloud-based services..

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Ms. K. Anuranjani, Assistant Professor, Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research, for her valuable guidance, continuous support, and insightful suggestions throughout the development of this project.

The authors also extend their appreciation to the Department of Computer Science and Engineering, BIHER, for providing the necessary resources and academic environment to successfully carry out this research work.

REFERENCES

[1] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surveys*, vol. 41, no. 3, pp. 1–58, July 2009.

[2] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153–1176, 2016.

[3] K. Yampolskiy and V. Govindaraju, "Behavioural biometrics: A survey and classification," *International Journal of Biometrics*, vol. 1, no. 1, pp. 81–113, 2008.

[4] [4] F. Chollet, *Deep Learning with Python*. Shelter Island, NY, USA: Manning Publications, 2018.

[5] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.

[6] [6] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[7] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[8] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, 1986.

[9] M. Ahmed, A. N. Mahmood, and J. Hu, "A survey of network anomaly detection techniques," *Journal of Network and Computer Applications*, vol. 60, pp. 19–31, Jan. 2016.

[10] S. Axelsson, "Intrusion detection systems: A survey and taxonomy," *Technical Report*, Chalmers University of Technology, 2000.