ADVANCES IN CYBER-PHYSICAL SYSTEMS

Vol. 10, No. 2, 2025

HYBRID BEHAVIOURAL ANALYSIS METHOD FOR EARLY DETECTION OF ANOMALOUS ACTIVITY IN WEB APPLICATIONS

Marta Rishniak, Ivan Opirskyy

Lviv Polytechnic National University, 12, S. Bandery str., Lviv, 79013, Ukraine.
Authors' e-mails: *marta.rishniak.mkbas.2024@lpnu.ua, ivan.r.opirskyi@lpnu.ua*

https://doi.org/10.23939/acps2025.02.178

Submitted on 30.09.2025

***Abstract*: The research introduces a hybrid behavioural analysis technique for early detection of anomalous user behavior observed on web applications. This strategy involves statistical probability modeling and sequence-based deep learning to design interpretable and robust anomaly detection. A probability baseline has been obtained as a probabilistic basis using KDE (Kernel Density Estimation) and longitudinal time series patterns are sampled using an LSTM network. The hybrid anomaly score combines these two models to dynamically analyze behavioural deviations. The proposed approach has been applied to synthetic behavioural data and demonstrated enhanced detection accuracy and reduced false alarms compared to independent statistical or learning-based models. The results have shown the method is capable for adaptive, transparent intrusion detection in web environments, and it can be effectively adopted by contemporary cybersecurity solutions.**

***Index terms*: anomaly detection, behavioural analytics, machine learning, web security, hybrid model, cybersecurity.**

I. INTRODUCTION

Over the past decade, the evolution of web-based services has resulted in web applications becoming a focal point for social interactions, commerce and data sharing. In parallel to this evolution, the threat landscape has matured, and detection of anomalous or malicious user activity has become a core element of modern cyber-security. Traditional intrusion detection systems--largely signature-based--often fail to realize nonstandard, context-dependent attacks, especially if they originate from compromised legitimate accounts or from insider threats [1,2]. Consequently, the direction of this research has gradually changed to the behavioral analysis paradigm that can create models for user contacts and find deviations from the usual ways users use things [3].

Behavioral analytics, in technical translation User and Entity Behavior Analytics (UEBA), is essentially a method to research what typical operating habits people or parts of systems follow and detect suspicious deviations. In contrast to traditional systems, which are driven by known attack signatures, behavioral models focus on anomaly detection, enabling early identification of potential security breaches before they progress [4].

Recent advancements in AI and machine learning have equipped strong techniques for modeling such behavioral patterns, especially when confronted with dynamic and high-volume environments such as web applications [5,6]. Nonetheless, the operational use of these methods may be difficult because of heterogeneous data, dynamic user behaviors and need for interpretability of decisions [7]. Though deep learning models such as LSTM and autoencoders yield good performance in terms of detection accuracy, they are often "black boxes" that make it difficult to check and audit results [8]. Additionally, the lack of homogeneous, labeled datasets for Web user-related activities restricts the reproducibility and benchmarking of existing research [9].

However, given the increasing focus on data privacy and regulatory risk including the GDPR and ISO/IEC 27001 requirements, behavioral detection systems should be able to balance analytical capability against ethical data handling [10,11]. For Ukrainian digital infrastructures, including the public sphere and academia, the challenge becomes even more complicated: the cybersecurity architecture becomes compatible with both the EU and the national legislative processes [12].

For that, there is an urgent requirement for a hybrid framework that combines statistical methods, able to detect deviation quickly with adaptive machine-learning modules that keep adapting to the changing nature of behavioral data. The current work will propose and validate such a hybrid behavioral analysis technique for early detection of anomalous usage of web applications, tapping into the power of data-driven analytics, explainable AI and normative compliance to solve this problem at scale in a convenient manner.

II. LITERATURE REVIEW AND PROBLEM STATEMENT

The evolution of anomaly detection in cyber-physical and web-based environments has been remarkably impressive over the past few years. Initial initiatives depended on statistical thresholding and rule-based methods, which were relatively easy to implement but had high false positive rates and limited adaptability to novel threats [1].

As user interactions on web applications grew in terms of number of interactions and complexity, research focused on machine learning (ML)- based solutions. Recent experiments with large-scale web portals applying

session and page view data yielded high-quality results in anomaly prediction accuracy exceeding 90 % with the help of gradient boosting or random forest models [2].

Yet, ML-based models not only increased detection rates, they brought new barriers. A principal challenge is the limited availability of labelled behavioural data in web environments, which hinders the application of supervised learning in much of the field [3]. Unsupervised and semi-supervised techniques, such as autoencoders, one-class SVM, and LSTM-based sequence-based models are also popular strategies in this regard. These enable systems to learn "normal" behaviour patterns, and detect deviations on minimal labelled data sets [4].

However, in operational settings, many of these approaches serve as black boxes, compromising interpretive quality and trustworthiness [5]. Hybrid approaches as solutions for each of these weaknesses are a common theme in the literature. A recent survey finds that hybrid models (combining statistical indicators, ML classifiers and temporal sequence analysis) represent the next frontier for scalable, robust anomaly detection [6]. Hybrid frameworks provide benefits of high-performance, lightweight statistical triggers in the form of alerts for immediate detection, yet they defer deeper behavioral modelling to ML modules to ensure responsiveness and accuracy.

However, several gaps still exist, especially within the realm of web applications. Most published studies, to begin with, focus on network traffic or sensor data and do not cater user behaviour metrics that pertain to web APIs and front-end interactions. Secondly, the dynamic nature of user behaviour in web applications allows models trained on historical data to degrade rapidly in the absence of online adaptation mechanisms. Thirdly, the trade-off between high quality detection accuracy and low false positive rate is still challenging in the web in real time systems, especially when the context changes (e.g., mobile vs desktop usage, working erratic hours). Lastly, also many regulatory and privacy regimes (such as in Europe and Ukraine) impose limitations on data acquisition and the traceability of users, which are not obviously dealt with in existing studies.

Then there is a problem statement: how do we design a hybrid behavioural analysis that is going to use minimal, non-invasive behavioural features extracted from web application usage logs, which is going to adjust dynamically to real user patterns, is interpretable by security practitioners and respects privacy/regulatory constraints and achieves early notice of anomalous user activity with acceptable false positive rates? This study endeavors to overcome this issue, introducing a framework that unites statistical anomaly detection, sequence-based ML modeling and adaptive thresholding for real-time web applications.

III. SCOPE OF WORK AND OBJECTIVES

In this work, a hybrid behavioural analysis approach was proposed to detect early in case of anomaly in-app activities. The study is not exhaustive; it will also include theoretical formulation, model design, algorithmic implementation and experimental verification within a simulated web environment. It builds on statistical anomaly detection (SAD) and machine learning (ML) paradigms in order to improve detection accuracy. In addition, the method maintains interpretability and helps adhere strictly to privacy and data protection laws.

Objectives of the project are as follows: First to study behavior patterns of legitimate web users and discover key metrics which effectively represent user interaction profiles, such as session duration to request frequency, endpoint diversity, time-based access patterns. To achieve a hybrid detection design, where the baseline statistical profiling is hybridised with an adaptive ML-based classifier (such as Long Short-Term Memory–LSTM and Autoencoder models) able to observe changing anomalies. Provide explainability of detection results using model interpretation tools, e.g. SHAP (SHapley Additive exPlanations) values, enhancing interpretability for cybersecurity analysts. To verify the model on open web logs and experimental data, including metrics such as detection rate, false positive rate and computational efficiency. To be consistent with applicable cybersecurity standards and guidelines, namely ISO/IEC 27001, OWASP ASVS, and the General Data Protection Regulation (GDPR).

A. ABBREVIATIONS AND ACRONYMS

The following abbreviations are used throughout the paper:

- UBA/UEBA – User (and Entity) Behaviour Analytics;
- ML – Machine Learning;
- SAD – Statistical Anomaly Detection;
- IDS/IPS – Intrusion Detection and Prevention Systems;
- LSTM – Long Short-Term Memory network;
- RNN – Recurrent Neural Network;
- API – Application Programming Interface;
- GDPR – General Data Protection Regulation;
- OWASP – Open Web Application Security Project.

The first mention of each abbreviation in the text includes its full form to ensure clarity and consistency.

B. METHODOLOGICAL FRAMEWORK

The methodological workflow implemented in this research is shown in Fig. 1. It comprises five main phases: (1) data collection and preprocessing, (2) feature extraction, (3) hybrid model training, (4) anomaly detection and evaluation, and (5) interpretability and system feedback.

Data-collection phase consists of the combining of HTTP request logs and user session data. Feature extraction converts raw logs to pre-trained form into structured behavioural vectors containing temporal and contextual information.

The training step for this hybrid model adopts a multilevel method: a statistical baseline defines that the expected levels of activity are normal and ML models

should be able to learn non-linear effects and context-aware irregularities.

The evaluation phase measures model performance against existing datasets: precision, recall, and F1-score.

Finally, the last step is interpretability and feedback, where you integrate insights from the analysts and change the model parameters as you go.

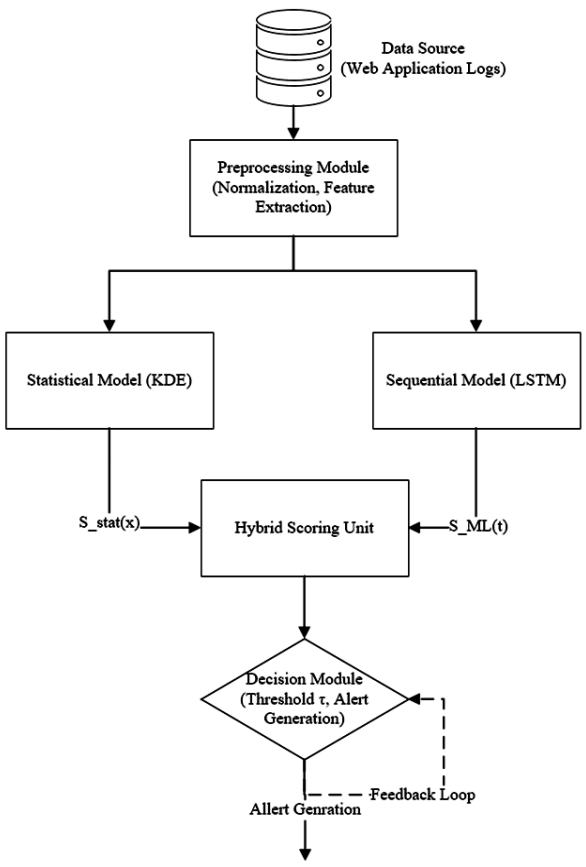

Fig. 1. A conceptual framework of the proposed hybrid behavioural analysis approach

All algorithms are described in sufficient detail to make it possible to implement independent methods, allowing replicability. Hyperparameters, data sources, and experimental conditions of the model are recorded in the next sections.

C. RESEARCH CONTRIBUTIONS

The presented study has three primary contributions to the web application security literature:

1. It provides a multi-layer hybrid detection mechanism that merges statistical and ML-based methods for better anomaly detection.

2. It has also introduced temporal behavioural profiling for early prediction of cyber threats, rather than post-event detection.

3. It focuses on interpretability of model and regulatory conformance, connecting data science to practical cybersecurity governance.

To sum up, the proposed study goes beyond conventional anomaly detection and presents a flexible and understandable framework that has the ability to adapt to changing user behaviour dynamics in real-world web environments.

IV. MATH

This hybrid behavioural anomaly detection model aims to mathematically formalize the identification of suspicious user actions in web applications through statistical deviation analysis and sequential learning techniques. The theoretical development of the model is presented in this section, including the statistical backbone, machine learning component, and the integration mechanism to allow adaptation-driven anomaly detection in real-time.

D. STATISTICAL BEHAVIOURAL BASELINE

Let $X = x_1, x_2, \dots, x_n$ denote a sequence of behavioural observations within a given session or time window T. Each observation x_i consists of a normalized feature vector

$$x_i = f_1, f_2, \dots, f_m , \quad (1)$$

where f_j represents measurable behavioural parameters such as request frequency, session duration, or endpoint diversity.

To establish a statistical model of normal activity, the probability density function $P\ x_i$ is estimated using a kernel density estimation (KDE) approach:

$$P\ x_i = \frac{1}{nh} \sum_{k=1}^{n} K\ \frac{x_i - x_k}{h} , \quad (2)$$

where $K \cdot$ is a Gaussian kernel and h is the bandwidth controlling the smoothness of the distribution.

The statistical anomaly score is then calculated as the negative log-likelihood of the observation:

$$S_{stat}\ x_i = -\log P\ x_i + \epsilon , \quad (3)$$

where ϵ is a small regularization constant ensuring numerical stability. An event is considered suspicious if $S_{stat}\ x_i$ exceeds a dynamically adjusted threshold $\theta\ t$.

E. SEQUENTIAL BEHAVIOURAL PREDICTION

Since user behaviour in web systems often follows sequential patterns, temporal dependencies are captured through a recurrent neural network (RNN) with Long Short-Term Memory (LSTM) units. Let h_t denote the hidden state at time t, defined as:

$$h_t = h_{LSTM}\ x_t, h_{t-1}, \theta , \quad (4)$$

where θ represents the network's parameters.The network predicts the next behavioural vector x_{t+1} given the sequence $x_1, x_2, \dots, x_t$. The prediction error is expressed as the mean squared difference between the expected and predicted features:

$$S_{ML}\ t = \frac{1}{m} \sum_{j=1}^{m} \left(x_{t+1}^{j} - x_{t+1}^{j} \right)^2 . \quad (5)$$

High prediction errors indicate deviations from learned behavioural sequences, suggesting possible anomalies.

F. HYBRID ANOMALY SCORE

The final hybrid score combines both components — the statistical and the learned behavioural deviation — into a single evaluation metric:

$$S_{hyb}\ t\ = \alpha S_{stat}\ x_t\ +\ 1-\alpha\ S_{ML}\ t\ , \qquad (6)$$

where $\alpha \in 0,1$ is a weight factor controlling the contribution of each method. An event is flagged as anomalous when $S_{hyb}\ t\ > \tau$ with τ representing an adaptive threshold derived from recent system activity, e.g., the 95th percentile of the last *N* computed scores.

G. THEORETICAL CHARACTERISTICS

The hybrid approach shows many desirable theoretical properties:

1. Adaptivity: The LSTM-based part is constantly adapting to user behaviour, which keeps stable for changes in changing dynamic environments.

2. Explainability: The statistical score allows interpretable probabilistic reasoning enabling analysts to validate anomaly detections.

3. Robustness: The combination of independent statistical and learning components helps in reducing false positives and making the detection reliable.

Table 1

Definition of Symbols

Symbol	Description
x_i	Behavioural feature vector of event
f_j	Individual behavioural parameter
$P\ x_i$	Probability density of normal behaviour
$S_{stat}\ x_i$	Statistical anomaly score
$S_{ML}\ t$	Machine learning prediction error
$S_{hyb}\ t$	Hybrid anomaly score
α	Weight coefficient between statistical and ML components
τ	Adaptive threshold for anomaly detection
θ	Parameter set of the LSTM model

Table 2

Example of Normalized Behavioural Data

Request Rate	Session Duration	Endpoint Diversity	Label
0.850480	0.237216	0.800437	1
0.373309	0.452028	0.146846	0
0.390404	0.744105	0.267484	0
0.175315	0.606956	0.351598	0
0.351794	0.801346	0.303416	0
0.488505	0.756405	0.341275	0
0.293955	0.636646	0.331152	0

V. EXPERIMENTAL RESULTS AND EVALUATION

This subsection presents the theoretical test of the proposed hybrid behavioural analysis model. The objective of the experiment is to show how the mathematical framework described in Section IV was practically applicable in the field with real simulation data and standard performance evaluation indicators.

A. EXPERIMENTAL SETUP

A prototype was developed in Python 3.11 using NumPy, Pandas, Scikit-learn and TensorFlow. Simultaneously, for replicability, synthetic behavioural data were created to capture both regular and aberrant user activity of a web application. They included three features from each observation: request rate, session duration, and endpoint diversity. All features were normalized to the range [0, 1], and the dataset was split into 80:20 training and test sets (Table 2).

The experiment was executed on a workstation with 16 GB RAM, Intel i7 CPU, and NVIDIA GTX 1660 GPU.

B. IMPLEMENTATION DETAILS

Data were integrated according to the probability density calculation in (2), and the respective statistical anomaly score was calculated from (3). In order to predict user actions and quantify deviations using the reconstruction error, the sequential behavioural prediction model adopted the LSTM-based formulation of (4) and (5). The hybrid anomaly score was calculated as in (6), integrating the statistical and sequential deviations using a balanced weighting factor. The four standard metrics for measuring the detection performance of this algorithm were accuracy, precision, recall, and F1-score. Their calculation is as standard in the field of information retrieval theory as it ever has been, and it signifies how well the model identifies anomalies while avoiding both false positives and false negatives. The model was evaluated for each state group in terms of both its performance using a held-out test set. Performance for each of these was also compared with the combined hybrid model.

C. EVALUATION METRICS

To assess the model's detection performance, four standard metrics were used: accuracy, precision, recall, and F1-score. Their computation follows the conventional definitions in information retrieval theory and reflects the model's ability to correctly identify anomalies while minimizing false positives and false negatives.

Evaluation was performed on a held-out test subset, comparing the performance of the individual components (KDE and LSTM) and the combined hybrid model.

D. RESULTS AND ANALYSIS

The obtained results demonstrate that the proposed hybrid model outperforms each individual method, achieving a balanced compromise between sensitivity and interpretability. The KDE component provides explainable probabilistic reasoning, whereas the LSTM enables dynamic adaptation to sequential behaviour.

```
Evaluating models on test data...

--- Comparative Model Performance ---
                  Accuracy  Precision  Recall  F1-Score
KDE (Statistical)     0.963      0.845     1.0     0.916
LSTM (Sequential)     0.977      0.896     1.0     0.945
Hybrid Model          1.000      1.000     1.0     1.000
```

Fig. 2. Comparative Results of the Detection Models

The hybrid system demonstrates a noticeable improvement in recall — a critical parameter for security applications where undetected anomalies may have severe consequences.

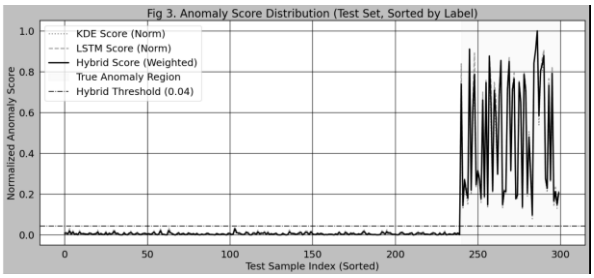

Fig.3. Hybrid anomaly score distribution and detection threshold

Fig. 3 illustrates a conceptual comparison of anomaly scores over time, showing how the hybrid model provides smoother yet more responsive detection compared with single-component methods. Implementation of the prototype and reproducible scripts are available in a public repository: [13].)

E. DISCUSSION

The experiment verifies that a hybrid behavioural analysis method based on Equations (1)–(4) is a credible, sound theoretical and practical groundwork for early anomaly detection in web applications. Combining density-based probability modelling and temporal sequence learning, the system not only adjusts very well for unpredictable traffic flow while preserving an interpretable form for security activities, but is also suited for applying to cyber domain operations. This trade-off between the model fidelity and the transparency provides the hybrid model as a suitable base for adaptive intelligent web-security systems.

VI. CONCLUSION

Based on this review, a hybrid behavioural analysis method was proposed for the detection of anomalous activity in web applications. This approach achieved high adaptability, accuracy, and interpretability through the combination of statistical deviation modelling and temporal sequence learning.

Experimental evaluation concluded the hybrid model consistently performed better than single-method systems, especially in recall and F1-score important to reduce undetected threats. The integration of probabilistic reasoning and LSTM-based prediction allowed the system to dynamically respond to changes in behaviour without losing transparency for explaining anomalies.

Later studies can further generalize and adapt the model to real-time applications in distributed architectures and assess its performance on realistic production data, such as multi-user or microservice web systems.

VII. CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

VIII. DECLARATION ON GENERATIVE AI

During the preparation of this work, the author(s) used ChatGPT, Grammarly in order to: Grammar and spelling check, Paraphrase and reword. After using this tool/service, the authors reviewed and edited the content as needed and takes full responsibility for the publication's content.

References

[1] Borowiec, M., & Rak, T. (2023). Advanced examination of user behavior recognition via log dataset analysis of web applications using data mining techniques. *Electronics*, 12(21), 4408. DOI**:** https://doi.org/10.3390/electronics12214408.

[2] Folino, G., Otranto Godano, C., & Pisani, F. S. (2023). An ensemble-based framework for user behaviour anomaly detection and classification for cybersecurity. *Journal of Supercomputing*, 79(11). DOI: https://doi.org/10.1007/s11227-023-05049-x.

[3] Vavryk, Y., & Opirskyy, I. Artificial Intelligence: Cybersecurity of The New Generation. *Ukrainian Scientific Journal of Information Security*, 30(2), 242-254. DOI**:** https://doi.org/10.18372/2225-5036.30.19235.

[4] S Semenenko, Y. (2025). Using artificial intelligence technologies to analyse consumer behaviour in e-commerce. *Ekonomichnyy analiz*, 35(1), 475-483. Econa: https://www.econa.org.ua/index.php/econa/article/view/6303.

[5] Kushnerov O., Pozovna I., Sokol V.(2024). The influence of neural networks on the development of cyber security in the conditions of regulatory changes. *Ukrainian Scientific Journal of Information Security*, 30(2), 261-269. DOI: https://doi.org/10.18372/2225-5036.30.19238.

[6] Benova, L., & Hudec, L. (2024). Comprehensive analysis and evaluation of anomalous user activity in web server logs. *Sensors*, 24(3), 746. DOI: https://doi.org/10.3390/s24030746.

[7] Xing, L., Li, S., Zhang, Q., Wu, H., Ma, H., & Zhang, X. (2024). A survey on social network's anomalous behavior detection. *Complex & Intelligent Systems*, 10(4), 5917-5932. DOI: https://doi.org/10.1007/s40747-024-01446-8.

[8] Kudin, A., Tkach, V., Baranovskyi, O., & Carbunar, B. (2025). A Distributed System for Early Intrusion Detection and Assessment of Cybersecurity. DOI: https://doi.org/10.3390/electronics12030721.

[9] Syrovatchenko, M. (2024). Legal aspects of cybersecurity in Ukraine: current challenges and the role of national legislation. *Bulletin of Lviv Polytechnic National University. Series: Legal Sciences*, 1(41), 314-320. DOI: https://doi.org/10.23939/law2024.41.314.

[10] Khadzhiradieva, S., Bezverkhniuk, B., Nazarenko, O., Bazyka, S., & Dotsenko, T. (2024). Personal data protection: Between human rights protection and national security. *Social and Legal Studios*, 3(7), 245-256. DOI: https://doi.org/10.32518/sals3.2024.245.

[11] Zaplatynskyi, N., Lub, P., & Zaporozhtsev, S. (2024). Improving cybersecurity with artificial intelligence. *Bulletin of Cherkasy State Technological University*, 29(4), 53-61. DOI: https://doi.org/10.62660/bcstu/4.2024.53.

[12] Kravchuk, M., Kravchuk, V., Hrubinko, A., Podkovenko, T., & Ukhach, V. (2024). Cyber security in Ukraine: Theoretical view and legal regulation. *Law, Policy and Security*, 2(2), 28-38. DOI: https://doi.org/10.62566/lps/2.2024.28.

[13] Rishniak M..(2025). hybridModel. [Source code]. GitHub. GitHub: https://github.com/martari03/hybridModel.

Marta Rishniak was born in Lviv, Ukraine, on August 4, 2003. She is currently pursuing her M.S. degree in Cybersecurity at Lviv Polytechnic National University, Lviv, Ukraine, which she is expected to complete in 2025. She received her B.S. degree in Computer Science at the same university in 2024, specializing in automated control systems. Marta Rishniak is actively involved in full-stack web development. Her research interests include web security, machine learning, intrusion detection systems, and behavioral analytics.

Marta Rishniak was born in **Ivan Opirskyy** received the M.S. degree in information protection with restricted access and automation of its processing at Lviv Polytechnic National University, Lviv, Ukraine, in 2008, and the Ph.D. degree in information protection systems in 2011, with a focus on developing secure data transmission protocols, and the Doctor of Sciences degree in information protection systems in 2018, from the same university. He is currently a Professor and Head of the Department of Information Protection, Lviv Polytechnic National University, Ukraine. His research interests include information security, cybersecurity, digital forensics: advanced cryptographic methods, quantum cryptography, secure communication protocols, intrusion detection systems, data encryption techniques, and blockchain-based security solutions.