IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.0429000

A Privacy-Preserving Behavioral Authentication Framework for Public Transport Applications

DAVID MONSCHEIN[1], MARKUS KNECHT[5], MUSTAFA AL SAMARA[2], JOSÉ ANTONIO PEREGRINA[1], TIM PIOTROWSKI[1], SASCHA GOHLKE[1] ABDELHAFID ABOUAISSA[2], WILFRID AZAN[3], ISMAIL BENNIS[2], ALFONSO GARCÍA DE PRADO[7], MARC GILG[2], EVA-MARIA NEUMANN[6], ZOLTAN NOCHTA[1], GUADALUPE ORTIZ[7], IOAN SZILAGYI[4], OLIVER P. WALDHORST[1], CHRISTIAN ZIRPINS[1]

[1]Karlsruhe University of Applied Sciences, Institute of Data-Centric Software Systems (IDSS), Moltkestr. 30, 76133 Karlsruhe, Germany
[2]Université de Haute Alsace (UHA), 2 rue des Frères Lumière, 68093 Mulhouse Cedex, France
[3]Université Lumière Lyon 2, Quai Claude Bernard, Lyon, Auvergne-Rhône-Alpes, France
[4]Neomia.ai, 3 Rue Pau-Henri Spaak, 68390 Sausheim, France
[5]Fachhochschule Nordwestschweiz (FHNW), Bahnhofstrasse 6, 5210 Windisch, Switzerland
[6]INIT GmbH, Käppelestraße 4-10, 76131 Karlsruhe, Germany
[7]Universidad de Cádiz (UCA), Av. Universidad de Cádiz, 10, 11519 Puerto Real, Cádiz

Corresponding author: David Monschein (e-mail: david.monschein@h-ka.de).

ABSTRACT Public transportation systems are increasingly relying on mobile services for ticketing, trip planning, and account management. In these systems, it is crucial to ensure secure and user-friendly authentication, as traditional login mechanisms often interrupt the travel experience and only provide protection during the initial login. Behavioral authentication based on interaction patterns and sensor data offers a promising solution for continuous verification. However, its practical deployment raises challenges related to privacy, robustness, distributed model training, and integration with existing authentication infrastructures. This paper presents the design and deployment of a privacy-preserving behavioral authentication framework for public transportation applications. The framework focuses on system-level integration of established privacy-preserving machine learning and authentication technologies. The framework combines client-side preprocessing and homomorphic encryption of behavioral data with server-side machine learning-based analysis. It incorporates model optimization techniques to improve robustness against impersonation attacks and supports collaborative model training across providers through federated learning with coordinated preprocessing and data governance. To enable practical deployment, the framework integrates behavioral authentication into existing OAuth 2.0 and OpenID Connect-based infrastructures, including authentication flows, token management, and cross-provider interoperability. We evaluated the proposed framework through preliminary evaluations based on a public dataset and a field test with a real-world public transport application, and a second field test assessing the feasibility of the federated learning setup in a collaborative training scenario. The results demonstrate the practical feasibility of the proposed architecture and provide insights into the challenges of deploying privacy-preserving behavioral authentication in real-world mobility ecosystems.

INDEX TERMS Behavioral Authentication, Behavioral Biometrics, Privacy-Preserving Machine Learning, Real-World Deployment, Federated Learning, Homomorphic Encryption, Authentication Infrastructure

I. INTRODUCTION

In mobile online services, such as e-commerce or public transportation applications, providing secure and reliable user authentication is crucial, as it guarantees that only authorized persons can access functionality and personal information. In the domain of public transportation, this requirement becomes even more critical. Mobile ticketing, subscription management, and personalized journey planning increasingly replace physical tickets. This places the smartphone at the center of the travel experience. Authentication in this context prevents misuse of services and safeguards user accounts against unauthorized access.

Among various approaches, *behavioral authentication* [1] has emerged as a promising solution. It allows continuous verification of identity without interrupting the user experience by leveraging unique patterns in interactions, such as

movement characteristics or touch dynamics [1]. In practice, behavioral authentication systems create a profile for each user based on their interaction patterns. The system then compares incoming behavioral data to the profile to verify the user's identity without requiring any explicit action. This aligns well with the environment of public transportation, where passengers expect security mechanisms to operate transparently in the background.

However, applying behavioral authentication to public transport applications faces several challenges. **First of all, behavioral authentication inherently raises privacy concerns.** A public transportation application typically follows the client–server model [2], where the application server must reliably verify the identity of the client user. When behavioral authentication is used in this setting, user privacy is a key problem: the service provider must process sensitive behavioral information, such as motion and interaction patterns, which can reveal personal habits and identifiable traits [3].

Second, authentication accuracy is highly context-dependent. The performance of behavioral authentication systems varies significantly depending on the operational environment. Variations in passenger behavior, device handling, or travel conditions can significantly influence performance, as shown in prior studies on mobile behavioral biometrics [1]. For instance, evaluations have shown that the effectiveness of a system can vary greatly depending on the application setting and potential attack scenarios [4].

Third, behavioral authentication requires sufficient amounts of realistic training data, which are difficult to obtain in distributed transport ecosystems. Training robust behavioral authentication models typically requires access to large, diverse datasets that capture user behavior across different environments. Ideally, these datasets consist of real data reflecting the target scenario [4]. In practice, however, this data is often distributed among multiple providers, and privacy regulations prevent direct sharing [5].

Finally, practical deployment requires seamless integration into existing authentication infrastructures. For real-world operation in public transport systems, the authentication results generated by a behavioral authentication framework must be integrated into existing authentication protocols used by transport providers. This is particularly relevant in cross-border scenarios or cooperative networks involving multiple service providers, where interoperability is essential.

Existing work on privacy-preserving behavioral authentication addresses important sub-problems of the overall challenge. Cryptographic techniques and data-filtering methods protect sensitive behavioral data during processing [3], [6], [7], while Federated Learning (FL) has been proposed to overcome data sharing restrictions in distributed environments [8], [9]. In parallel, prior research has investigated behavioral biometric models and optimization strategies to improve robustness against attack scenarios [10], [11], [12]. However, these contributions are considered in isolation and are not combined into a unified, deployable framework. In particular, the coordinated integration of privacy-preserving data processing, collaborative model training with common preprocessing standards and data governance mechanisms, and seamless embedding into established authentication protocols remains unexplored. Moreover, once models are deployed, many existing solutions do not address their integration into established authentication flows, which complicates adoption in real-world applications. This gap becomes especially evident in cross-provider and cross-border public transport scenarios, where multiple service providers must collaborate under strict privacy, regulatory, and operational constraints.

To address ongoing challenges, the *aura.ai* research project[1] develops a privacy-preserving behavioral authentication framework to support the entire lifecycle of behavioral authentication in a cross-provider and cross-border setting. The framework combines several established techniques into a unified architecture tailored to public transportation applications. Behavioral data is protected using Homomorphic Encryption (HE) on the client side before transmission, enabling Machine Learning (ML)–based analysis without exposing sensitive user information [13], [7]. We further employ model optimization techniques to improve the accuracy of behavioral analysis, especially when it comes to detecting complex attack patterns [12], [11]. Federated learning is incorporated as a component that allows multiple providers to collaboratively train models without disclosing their local datasets. This is complemented by collaboration mechanisms driven by data governance and coordinated preprocessing to ensure consistent model quality [14], [9], [8]. To achieve real-world applicability, the framework seamlessly integrates with existing authentication protocols and standards, thereby extending them to support behavioral authentication. This includes establishing authentication flows, managing token lifetimes, and decision thresholds, as well as enabling multi-provider interoperability. The latter supports cross-provider scenarios, also across national borders, while maintaining strict privacy boundaries between providers.

In summary, this work presents a privacy-preserving behavioral authentication framework tailored to public transportation applications, with a focus on the integration of existing techniques for privacy protection and ML in a real-world setting. The key contributions of this paper are:

1) We design and implement a framework for privacy-preserving behavioral authentication in public transportation applications that integrates multiple established techniques into a coherent authentication architecture.
2) We demonstrate how privacy-preserving behavioral authentication using homomorphic encryption can be incorporated into mobile applications while ensuring authentication accuracy and efficiency.
3) We integrate federated learning to enable multiple providers to collaboratively train models without sharing raw data, supported by coordinated preprocessing and collaboration mechanisms driven by data governance.

[1] https://www.h-ka.de/en/idss/aura-ai

4) We bridge the gap between existing authentication standards such as OAuth 2.0 and behavioral authentication by incorporating token lifetime management, decision thresholds, and support for cross-provider interoperability.

We validated our framework through a preliminary evaluation and two field tests. In the preliminary evaluation, we conducted experiments to evaluate specific components of the system under controlled conditions. More specifically, we measured the authentication performance on multiple public datasets to verify that the ML models achieve robust detection of attacker scenarios. In the first fest field test, we integrated behavioral authentication mechanisms into the *regiomove*[2] mobile application that is used for public transportation in Karlsruhe, Germany. Over the course of one month, selected participants regularly used the application. Additionally, attack scenarios in which an attacker attempted to imitate the behavior of a legitimate user were conducted. The results demonstrate that our approach is able to achieve accurate authentication decisions, while maintaining the efficiency necessary for seamless integration into real-world public transport applications, where users expect a continuous and unobtrusive experience. Nevertheless, we also identified several pitfalls that can cause wrong authentication decisions and led to reduced accuracy in some phases of the field test. With the second field test, we evaluated the feasibility of the FL setup within our framework. We performed a experiment where we request a set of students, who play the role of the different mobility providers, to negotiate over the FL training settings. Results showed that our platform was easy to use for the students, who managed to run a training in a maximum of two hours. Based on these observations, we derived valuable insights that can guide future experiments in the field of behavioral authentication and FL.

The paper is structured as follows. First, Section II introduces the scenario of public transport applications in detail, emphasizes the requirements for appropriate authentication solutions and outlines the threat model considered in this work. Next, Section III reviews prior work in the key areas that underpin our framework. Section IV outlines our core concepts and building blocks that make up the proposed framework. Afterwards, Section V presents a preliminary evaluation of the proposed approach using a public dataset under controlled conditions. Section VI then describes the field test conducted with the *regiomove* public transport application and reports the corresponding results. Section VII presents a second field test evaluating the feasibility of the federated learning setup in a collaborative training scenario. Finally, Section VIII summarizes key aspects of the paper and provides an outlook on future work.

II. MOBILE APPLICATIONS FOR PUBLIC TRANSPORT

In app-based public transportation systems, authentication must function reliably not only at login but also throughout the entire journey. Solutions need to work without interrupting the user, adapt to changing operational contexts, and comply with strict privacy regulations. This is especially important in environments where services are offered by multiple independent providers and passengers may switch between them during a single trip. In such cases, interoperability and trust across provider boundaries are essential, even when direct sharing of personal or behavioral data is not possible.

A. PUBLIC TRANSPORT APPLICATION SCENARIO

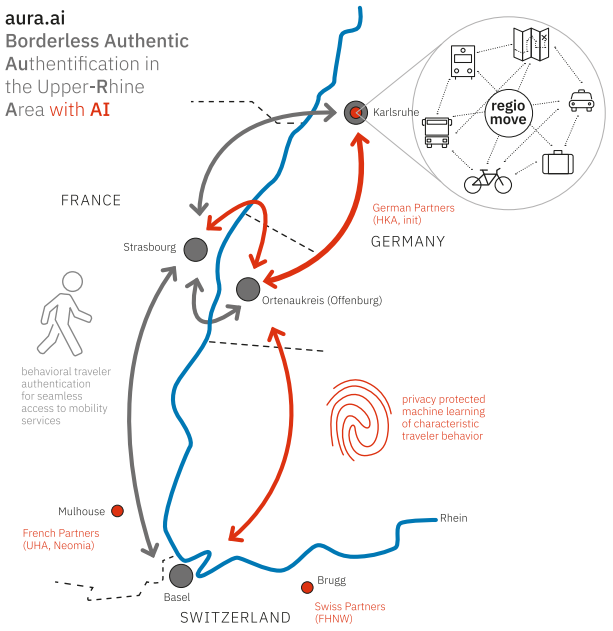


FIGURE 1: Example of a cross-border public transport scenario in the Upper-Rhine region

An illustrative example of such a setting is the tri-national Upper-Rhine region, shown in Figure 1, which depicts the core scenario addressed by the *aura.ai* research project. Spanning approximately 350 km along the borders of France, Germany, and Switzerland, the region is characterized by frequent cross-border commuting and a growing demand for integrated mobility services. Platforms, such as *regiomove* enable multimodal travel within a local association of providers. In particular, the *regiomove* application offers different services of local transport such as buses, trams, regional and urban trains, as well as car and bike sharing (cf. Section II). It is operated by *KVV*[3] and developed collectively by *raumobil*[4] and *INIT*[5]. *INIT* is responsible for the backend and identity provider. The backend contains the user management, booking platform, payment provider and communication with other service providers. *raumobil* manages the distribution of the application and the frontend. Beyond *regiomove*, another platform built on the same underlying infrastructure

[2]https://www.kvv.de/mobilitaet/regiomove.html

[3]https://www.kvv.de/

[4]https://www.raumobil.com/

[5]https://www.initse.com/

is *OrtenauMobil*[6]. The main difference is the operation by *Ortenaukreis*[7] and that it has its own service providers.

B. SYSTEM REQUIREMENTS AND CHALLENGES

In this environment, traditional approaches become impractical: operators are subject to different legal frameworks, preventing direct data sharing, and users expect a seamless travel experience without repeated manual logins. The main challenges in this scenario can be summarized as follows:

- **Privacy:** Behavioral authentication relies on sensitive motion and interaction data, which must remain protected.
- **Interoperability:** Authentication solutions must integrate with established protocols and standards while remaining functional in heterogeneous, multi-provider environments.
- **Adaptability:** Travel conditions, device handling, and passenger behavior vary widely, requiring models that remain accurate across contexts.
- **Data diversity:** Robust authentication requires behavioral data from multiple providers to capture the variability of user behavior across regions, yet legal and privacy constraints prevent the direct sharing of such data.

These factors define the operational and technical context for our proposed privacy-preserving behavioral authentication framework and inform the design choices discussed in this work.

C. THREAT MODEL SUMMARY

To clarify the security assumptions of the considered deployment scenario, we outline an informal threat model reflecting risks in mobile public transport applications.

- **Attack scenarios.** A primary threat arises when an attacker gains access to the legitimate user's device, for example if a device is stolen or left unattended during a journey. In such situations, the attacker may attempt to access tickets, perform bookings, or use stored credentials. Behavioral authentication is intended to detect such deviations in interaction patterns and identify illegitimate use.
- **Client-side adversaries.** Because of the potential attack scenarios, we assume that client devices cannot be trusted. A malicious or compromised client may deviate arbitrarily from the authentication protocol to imitate other users or gather behavioral information. The proposed framework mitigates such scenarios by building on a privacy-preserving authentication protocol that combines robust ML-based attack detection with provable guarantees for data confidentiality [7].
- **Server-side adversaries.** Service providers process authentication requests and behavioral data as part of the authentication process. However, servers may act maliciously, deviating from the authentication protocol in order to recover sensitive details from the behavioral information sent by clients. In particular, the system is designed to prevent service providers from accessing raw behavioral data that could reveal sensitive user characteristics or habits [7]. Privacy-preserving techniques such as HE are used to avoid the exposure of sensitive information during processing.
- **Trust assumptions.** The framework assumes that either the client or the server may behave maliciously, but not both simultaneously. Furthermore, providers participating in collaborative model training are assumed to be *honest-but-curious* [14] meaning that they follow the agreed training protocol but may attempt to infer additional information from the data they observe. Their local datasets remain private and are not shared directly with other participants.

These assumptions define the trust boundaries and attacker capabilities considered in this work and guide the design of the proposed privacy-preserving behavioral authentication framework. Formal security analyses of the individual building blocks used in the framework have been conducted in corresponding prior work [7], [14], [9]. In this paper, we focus on their system-level integration and deployment within the described threat model.

III. RELATED WORK

This section reviews existing research in four key areas that underpin our work: robustness in behavioral biometrics, privacy-preserving behavioral authentication, federated learning for distributed training, and integration with established authentication protocols.

A. ROBUSTNESS IN BEHAVIORAL AUTHENTICATION

Robustness is a critical requirement for behavioral biometric systems deployed in continuous authentication settings, particularly in real-world public transport applications exposed to diverse environmental conditions and adversarial behavior (cf. Section II).

With the increasing adoption of deep learning, numerous approaches have been proposed to improve recognition accuracy and robustness in behavioral authentication. For example, Deb et al. [15] proposed a continuous authentication system that combines interaction patterns such as keystroke dynamics with motion sensor data. Their approach uses a Siamese RNN-LSTM architecture trained with contrastive loss to learn discriminative user embeddings from heterogeneous sensor inputs. Similarly, Abuhamad et al. [16] introduced AutoSen, a behavioral biometric framework that leverages background sensors such as the accelerometer, gyroscope, and magnetometer, applying RNN-LSTM models with triplet loss to continuously authenticate users based on passive interaction data.

Other research has explored touchscreen-based behavioral biometrics. Acien et al. [17] investigated authentication based on swiping gestures and keystroke dynamics collected from smartphone interactions. Using Siamese neural networks and

[6]https://www.ortenaukreis.de/ortenau-mobil
[7]https://www.ortenaukreis.de

various loss functions, their work demonstrated that combining multiple behavioral modalities can improve authentication performance. Score-level fusion strategies across modalities were shown to reduce error rates compared to single-modality approaches.

To support benchmarking and comparative evaluation, Stragapede et al. [10] introduced the BehavePassDB dataset, which captures natural human–mobile interactions across multiple tasks and sensor modalities. Their work highlights the importance of evaluating behavioral authentication systems under different impostor scenarios, including both random and skilled impersonation attacks.

While these approaches demonstrate the potential of modern ML techniques to improve robustness in behavioral authentication, they typically assume centralized training and evaluation settings. As a result, they do not address challenges related to privacy-preserving data processing, distributed model training, or deployment in heterogeneous multi-provider environments.

B. PRIVACY-PRESERVING BEHAVIORAL AUTHENTICATION

Behavioral authentication enables continuous verification based on user interaction patterns, yet processing such data in client-server architectures introduces significant privacy concerns. Existing research can be broadly grouped into approaches based on Homomorphic Encryption (HE), data transformation, and alternative cryptographic protocols. HE-based solutions allow computation on encrypted data and thus offer strong privacy guarantees, but often suffer from high computational costs, latency, or restrictions in analytical capabilities [18], [3], [19], [20]. Data transformation techniques, such as cancelable biometrics or feature obfuscation, are more lightweight, but typically reduce accuracy and provide weaker protection [21], [22], [23]. Other strategies relying on other cryptographic methods, including private set intersection, biometric key generation, or format-preserving encryption, can be efficient or accurate in specific scenarios but may have limited applicability or residual privacy risks [24], [25]. Trade-offs between privacy, efficiency, and robustness remain across all categories, and many solutions lack integration with ML techniques and established authentication infrastructures.

Related work in image and video processing has explored privacy-preserving techniques such as watermarking to protect sensitive multimedia data during transmission, particularly in telemedicine settings [26], [27]. While these methods focus on ensuring the integrity and confidentiality of visual data, behavioral authentication addresses a different problem by analyzing user interaction patterns rather than protecting transmitted media content.

C. FEDERATED LEARNING WITH DATA GOVERNANCE

Federated Learning (FL) was conceived by McMahan et al. [28], who used the method to train the Google Keyboard on Android (GBoard) ML model for predictive text. After its conception, a whole field of FL research has appeared. Among the plenitude of FL research publications, those that fall close to our focus are collaboration means for FL, and security within the FL process.

Research in FL rarely focuses on the process preceding the training. Establishing different aspects, like the goals of the training, the data to be used, or the model to use for training, is needed before starting the FL process. For this reason, some negotiation and collaboration mechanisms must be put in place. The two main works dedicated to implementing collaboration platforms are Schelegel et al. [29] and FLIP [30]. Both platforms offer functionality to upload the FL code for training, as well as to connect all clients for training. However, only Schelegel et al. offer functionality for provenance, and none offer functionality for estimating how much a client is contributing to the training. Furthermore, the discussion around the data structure is not included in any of the platforms, being an essential part of the setup in FL.

In FL settings, model weights or gradients are sent to a central server for aggregation. Recent research showed that, based on a trained model, information about training data or even entire data items can be reconstructed. Related research in collaborative learning proves that these attacks can be applied in FL using the exchanged weights or gradients [31], [32], [33]. Therefore, the server presents a third party that must be trusted by all participants of the federated training. Using a decentralized implementation [34], the need for a central server can be removed. Nevertheless, peer group members can potentially attack other participants. In fact, the attack surface increases without the application of additional security mechanisms, as shown in related work [35].

For tabular data specifically, only a handful of works currently demonstrate practical reconstruction attacks [33], yet these already constitute a tangible risk to participants' data. Typical defenses include Differential Privacy and Secure Multi-Party Computation, which can prevent or substantially hinder reconstruction but often at nontrivial utility or system cost. An aspect that is frequently overlooked in this context is the preprocessing pipeline. Before training, participants transform raw data into model-ready representations and often need to exchange auxiliary statistics or artifacts—such as category vocabularies for one-hot encodings, scaling parameters, quantization thresholds, or clustering prototypes—that may themselves reveal sensitive information. Prior work implies that providing such preprocessing artifacts to a server can materially aid the reconstruction of clients' local data [33]. These observations motivate hybrid designs in which model training remains centralized while preprocessing is executed in a decentralized manner between the participants, thereby reducing the leakage surface by providing the server with private information, without incurring the full cost of end-to-end cryptographic training.

D. CROSS-PROVIDER AUTHENTICATION INFRASTRUCTURE

Behavioral data offers a valuable signal for detecting changes in user identity, but a complete authentication system also

requires identification, credential management, session handling, and access control. These aspects are already addressed by established protocols. Accordingly, new authentication mechanisms aiming for practical viability often integrate with established protocols such as OAuth 2.0 [36] or OpenID Connect [37].

Since both protocols traditionally involve the user directly, they lack support for background mechanisms. The emerging OAuth 2.0 for First-Party Applications extension [38] addresses this by introducing authorization challenges, which allow user-independent workflows suitable for behavioral authentication.

Cross-provider scenarios further highlight interoperability requirements. OpenID Connect was designed with such interoperability in mind, whereas OAuth 2.0 was not. To support OAuth-only providers, systems can rely on the OAuth 2.0 Token Exchange standard [39], which enables tokens issued through behavioral authentication to be exchanged for tokens accepted by individual providers.

Existing approaches to integrate behavioral authentication fall into one of two categories. Either they do not involve an authentication server and operate exclusively on the client device to perform local authentication [40], [41], or they rely on a separate monitoring component that observes user behavior during an active session and contacts the authentication server to terminate sessions or revoke tokens if misbehavior is detected [42]. Both approaches do not allow to use the behavioral data as a authentication factor on an authentication server.

IV. AUTHENTICATION FRAMEWORK

This section presents the design of the proposed privacy-preserving behavioral authentication framework, detailing its main components: the authentication mechanism, model optimization strategies, federated learning setup, and supporting infrastructure.

A. AUTHENTICATION FRAMEWORK OVERVIEW

Figure 2 summarizes the overall architecture of the proposed authentication framework and highlights the interaction between its main components. The framework enables privacy-preserving behavioral authentication across multiple mobility providers while integrating with existing authentication infrastructures.

Behavioral data is collected on the user's device during normal app usage. To preserve privacy, these interaction features are immediately encrypted using HE before being transmitted to the behavioral analysis service (CA Server). This component evaluates the encrypted input using neural networks trained for behavioral authentication, producing encrypted distance scores that reflect the similarity of the current interaction pattern to the user's established profile. After decryption on the client, the results are returned to the analysis service together with a validity proof ensuring correct decryption. The CA Server forwards them to the authentication server (CoNym) for storage and later use. This

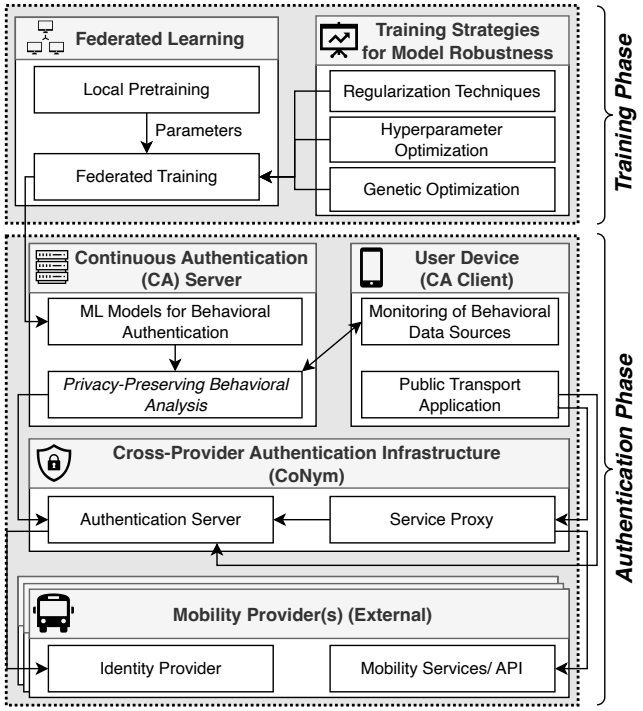


FIGURE 2: Architecture of the proposed privacy-preserving behavioral authentication framework and the interaction between its main components

process, detailed in Section IV-C, ensures that sensitive data remain confidential throughout the analysis pipeline.

The models employed by the analysis service originate from a collaborative federated learning setup (Section IV-E). In this setting, regional mobility providers train shared neural networks under a common data governance framework [9]. Each provider prepossesses its local datasets using privacy-preserving encoders and contributes only parameter updates to the global model. This approach enables knowledge sharing without the exchange of raw behavioral data and ensures that the deployed model remains consistent across organizational boundaries.

While training defines the analytical capability of the framework, robustness in behavioral biometrics (Section IV-B) ensures that these models remain reliable under real-world conditions. Techniques such as model regularization, hyperparameter optimization, and genetic optimization strengthen the resistance of the behavioral model against evasion and impersonation attempts. As a result, the distance scores delivered to the authentication infrastructure remain meaningful even under noisy or adversarial input.

The behavioral results stored by the authentication server are integrated into standard OAuth 2.0 workflows. When a client device requests authentication, the authentication server retrieves the most recently stored behavioral result corresponding to that device. If the result exceeds a configured threshold, the user is re-authenticated. Otherwise, authentication is delegated to an existing Identity Provider, typically

operated by the same organization as the backend services.

Requests to mobility services are likewise routed through the custom authentication server. If the authentication information in the request is valid it uses the OAuth 2.0 Token Exchange protocol to replace the authentication information with the information expected by the targeted service. This mechanism bridges the access control models of the behavioral authentication system and existing provider infrastructures, allowing services to operate unchanged while seamlessly benefiting from the new authentication layer.

The framework establishes a complete and practical pipeline for secure behavioral authentication in distributed mobility ecosystems through the interaction of these components: federated learning for model creation, robustness optimization for security, privacy-preserving analysis for user trust, and cross-provider integration for interoperability.

In the following, we first focus on the robustness and reliability of ML models targeting behavioral biometrics. These models are the analytical foundation of the privacy-preserving authentication protocol we are using.

B. ROBUSTNESS IN BEHAVIORAL BIOMETRICS

As a foundation for privacy-aware authentication, the behavioral models must be robust and reliable across real-world conditions. In mobile and public transport environments, authentication systems must resist a wide range of attack strategies, including random and skilled impersonation, behavioral data extraction, and poisoning or spoofing attempts that exploit device or contextual variability. Robustness in this context refers to the model's ability to maintain high discriminative performance and stability despite environmental noise, heterogeneous sensors, or deliberate adversarial manipulation.

This subsection summarizes the general principles and methodologies derived from earlier research contributions (B2CAR [12] and AI-MBBCA [11]), which address the fundamental challenge of making behavioral biometrics reliable under real-world variability. These models provide the analytical basis upon which the privacy-preserving version introduced in the next subsection is constructed.

Our earlier work, B2CAR [12], demonstrated that incorporating regularization techniques (Ridge, Lasso, ElasticNet, and Bayesian) into LSTM-based models significantly enhances resilience against both random and skilled impostor attacks. Regularization constrains model complexity, reduces weight variance, and mitigates overfitting to device-specific artifacts, thereby improving temporal generalization across users and sessions.

To further strengthen resilience, our AI-MBBCA framework [11] combined hyperparameter optimization using GA with an IF anomaly detector as a second-layer defense. This dual approach not only maximizes the accuracy of LSTM models trained with triplet loss but also introduces adaptive mechanisms to detect abnormal behavioral embeddings that may result from adversarial manipulation. Triplet loss promotes the formation of embeddings that place similar behaviors closer together and push embeddings of dissimilar or anomalous behaviors farther apart [10]. Experiments on the BehavePassDB dataset confirmed that this architecture consistently outperformed baseline models under complex attack scenarios, including attempts to mimic legitimate users' behavioral patterns. Together, these contributions highlight the importance of integrating both proactive model optimization and reactive anomaly detection for attack-resilient behavioral authentication in public transport applications. Figure 3 illustrates the AI-MBBCA framework architecture, an extension of the B2CAR approach that incorporates GA optimization and Isolation Forest (IF) for impostor detection. The framework employs a hybrid LSTM network trained with a triplet-loss function and enhanced by regularization to capture both spatial and temporal behavioral patterns, as shown in Figure 3. The GA automatically tunes critical hyperparameters (e.g., learning rate, batch size, margin, dropout rate), while the IF layer analyses the feature embedding space to detect anomalies indicative of adversarial or impersonation behavior. This multi-stage design improves adaptability and robustness against complex attack scenarios, providing a resilient foundation for continuous behavioral authentication in mobile environments.

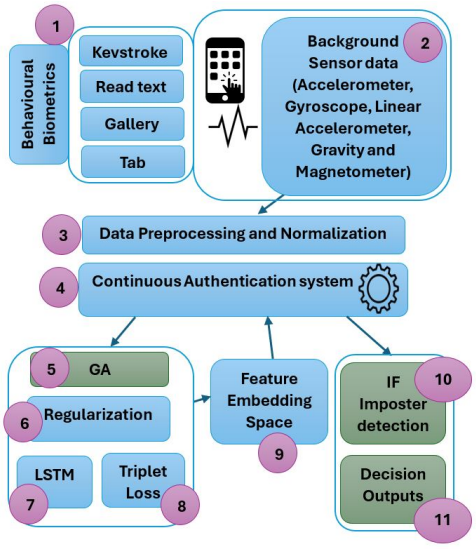


FIGURE 3: AI-MBBCA approach.

1) **Behavioral Biometrics:** Capture user-specific interaction patterns such as typing rhythm (keystroke), reading and scrolling behavior (read text), image navigation habits (gallery), and touch or swipe gestures (tab).
2) **Background Sensors:** Include accelerometer, gyroscope, linear accelerometer, gravity, and magnetometer signals that describe device motion and orientation. These continuously provide contextual behavioral data.
3) **Data Preprocessing and Normalization:** Raw sensor streams are cleaned, segmented, and normalized to ensure consistency and reduce noise before entering the learning pipeline.

4) **Continuous Authentication (CA) System:** Serves as the central unit coordinating data collection, feature extraction, and authentication through integrated modules such as LSTM, GA, and IF.
5) **Genetic Algorithm (GA):** Optimizes key hyperparameters (e.g., learning rate, dropout, batch size, LSTM units, regularization type, and triplet strategy) using evolutionary operations of selection, crossover, and mutation to enhance model generalization.
6) **Regularization:** Introduces penalty terms to the model loss to prevent overfitting and improve the robustness of the learned behavioral representation.
7) **LSTM:** Processes sequential sensor data to capture temporal dependencies and dynamic variations in user interactions.
8) **Triplet Loss:** Structures the feature space by minimizing the distance between genuine samples while maximizing separation from impostor samples, enforcing strong inter-class discrimination.
9) **Feature Embedding Space:** Represents users through compact feature vectors where genuine user embeddings form tight clusters and impostor embeddings remain distant, facilitating accurate decision making.
10) **Isolation Forest (IF) for Impostor Detection:** An unsupervised anomaly detection method that isolates abnormal behavioral patterns and distinguishes genuine users from impostors using learned embedding distributions.
11) **Decision Outputs:** Final authentication is derived by combining distance-based similarity scores and anomaly indicators, producing a robust, continuous verification decision in real time.

The evolution from static regularization (B2CAR) to adaptive optimization (AI-MBBCA) defines a three-layer robustness paradigm:

- **Preventive hardening:** Regularization constrains model capacity and mitigates overfitting, ensuring stable temporal representations.
- **Adaptive resilience:** GA-driven optimization enables dynamic model adaptation under variable behavioral or environmental conditions.
- **Reactive detection:** The IF layer identifies deviations in the embedding space, signaling potential adversarial or anomalous activity.

These robustness mechanisms complement the HE-based privacy preservation and the FL-based collaborative training introduced in the following subsections. Together they outline a secure and reliable end-to-end authentication pipeline. The robustness oriented approaches described in B2CAR [12] and AI-MBBCA [11] represent analytical extensions that can be incorporated into the privacy preserving architecture to improve resilience against sophisticated attacks. Accordingly, the privacy-preserving framework presented next adopts the robust modeling paradigm as its analytical core.

C. PRIVACY-PRESERVING BEHAVIORAL AUTHENTICATION

Building upon the robust behavioral models described in the previous subsection, we incorporate privacy-preserving mechanisms to ensure that the authentication process remains confidential and verifiable even in untrusted environments. We rely on an authentication protocol proposed and evaluated in prior work, which uses HE to protect user privacy [13], [7]. In particular, sensitive behavioral data is encrypted on the client device before being transmitted to the server, which processes data only in encrypted form. This process is illustrated in Figure 4.

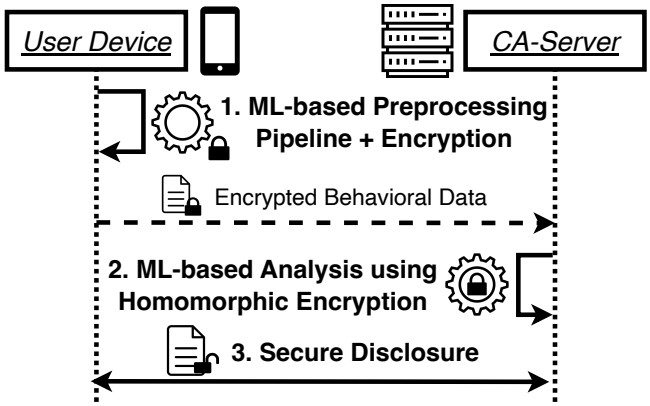


FIGURE 4: Excerpt from the privacy-preserving authentication process involving the User Device and the CA Server [13], [7]

The approach assumes an initial enrollment phase using conventional multi-factor authentication to establish a behavioral baseline. It is important to note that this baseline is also encrypted with HE [13]. The enrollment is followed by periodic re-authentication at fixed intervals (e.g., every 60 seconds) based on behavioral data. Standard security mechanisms such as encrypted communication and server authentication are employed to ensure the confidentiality and integrity of all exchanged messages [7].

Based on this setting, each behavioral authentication process is structured into three parts.

a: ML-based Preprocessing Pipeline

First, we require a monitoring component on the user device. This component collects data from different sensors that reflect behavioral characteristics. Common data sources that are suitable for behavior-based authentication include accelerometers, gyroscopes, magnetometers, and touchscreen inputs [1]. Then, upon an authentication, the collected data is first preprocessed and features are extracted from the raw data for each data source. The features represent measurable properties of the signal that capture behavioral patterns relevant for distinguishing users [43]. This means, for example, that the raw sensor data is aggregated over a certain period of time in order to align different sampling rates [13].

Afterwards, the extracted features from each data source are used as input for a separate neural network. The goal

of these neural networks is to reduce the data to a vector of fixed size (*embedding*) to save network traffic and computing power in the downstream analysis [13]. For each of the neural networks, the RNN architecture and the triplet loss training strategy introduced in Section IV-B is employed. Particularly, our framework is flexible and also allows the integration of other model architectures. The goal is to enforce that embeddings of the same user lie close together while those of different users are well separated, which is suited for authentication scenarios. Overall, the data collection, the feature extraction and the mapping to embeddings form our preprocessing pipeline. Following the execution of the pipeline, the output embeddings are encrypted. For this purpose, the Cheon-Kim-Kim-Song (CKKS) HE scheme is used, which enables efficient calculations on encrypted data [44]. Finally, the encrypted embeddings are transferred to the continuous authentication server (CA Server) as shown in Figure 2.

b: ML-based Analysis of Encrypted Data

The CA Server first retrieves encrypted reference embeddings from its database, where it stores baseline data for each user. As mentioned before, this baseline data is established during the enrollment phase, which uses conventional authentication before switching to behavioral biometrics [13], [7]. In this way, the CA Server ensures that the reference embeddings represent legitimate user behavior. In the next step, the received encrypted embeddings and the reference embeddings are entered together into a ML-based analysis. The ML-based analysis relies on a convolutional neural network (CNN) to determine whether the current behavior matches the behavior known from the legitimate user [7]. We can conduct this analysis due to the characteristics of the CKKS scheme, which allows to perform ML model inferences on encrypted data. The original work [7] proposes to analyze the embeddings of all data sources together with a comprehensive neural network structure that consists of multiple convolutional and dense layers. However, we chose to use smaller neural networks and analyze each data source's embeddings individually. The idea is to enhance efficiency and achieve shorter response times as HE tends to be computationally expensive [6]. In particular, each neural network consists of a convolutional layer, followed by an average pooling layer and two dense layers. After performing a neural network inference for each data source, we calculate the mean to come to a final score that reflects whether the embeddings originate from the same user or not. To train these neural networks, we sample embeddings from different user sessions, pairing embeddings from the same user as positive examples and embeddings from different users as negative examples, enabling the networks to learn to distinguish between legitimate and illegitimate behavioral patterns.

Once the average of all neural network inferences has been computed, the ML-based analysis of the encrypted embeddings is finished. However, the calculated score is still encrypted and can not be interpreted by the CA Server. For this reason, we rely on an exchange with the client device to securely disclose the analysis result to the server [7].

c: Secure Disclosure

The last part of the process is responsible for revealing the computed authentication result to the server while protecting integrity and confidentiality. In other words, the disclosure needs to make sure that the server learns the correct analysis result and that no behavioral data is leaked to the client or the server besides the actual authentication result. To achieve this, the integrated protocol relies on a Zero-Knowledge Proof (ZKP) [45]. ZKPs allow a prover to convince a verifier that a statement is true without revealing any additional information beyond the statement's validity [45].

As we only need to reveal a single ciphertext, the ZKP can be realized quite efficiently. In the ZKP process, the encrypted authentication decision is transferred to the client. The client then decrypts it using its secret key and generates a proof that the decrypted value corresponds to a claimed authentication result without revealing any additional information about the underlying secret key. For this purpose, a zk-SNARK circuit is used [7]. The reason for choosing zk-SNARK are small proof sizes, low verification times and the mature ecosystem that facilitates deployment on mobile devices [46]. More specifically, zk-SNARK constructions requires initial ceremony in which public parameters are generated. In our implementation, the CA Server performs this setup once and derives the corresponding proving and verification keys. The server has a strong incentive to execute this setup correctly, since any leakage of the secret randomness would allow malicious clients to forge proofs. If the server is not trusted to perform the setup alone, it can alternatively be realized through a multi-party computation ceremony involving multiple independent participants [7]. After generating the proof, the client sends it, together with the claimed authentication result, to the server. Lastly, the server verifies the proof and, if valid, passes the disclosed result to cross provider infrastructure called CoNym, that is introduced in the upcoming Section IV-D.

Overall, strong privacy protection is ensured by the fact that the server has no access to raw behavioral data. Moreover, potential attacks are mitigated by the protocol used, as has already been demonstrated in existing work [7]. In particular, the approach used is robust against malicious clients. A compromised client cannot manipulate the authentication outcome or extract behavioral information about other users by exploiting the protocol. This is important in the case of a mobile application for public transportation because attackers may gain access to a legitimate user's device or manipulate the client application in order to bypass authentication mechanisms. Additionally, the use of HE does not lead to a significant decrease in accuracy compared to alternative methods that remove sensitive features or apply alternative privacy-preserving techniques [7].

As training directly on encrypted data is currently not computationally feasible [47], the ML models used in both the preprocessing pipeline and encrypted analysis are trained

on unencrypted data. This can be done, for example, using public datasets, as our prior results indicate that the models generalize well to unseen users [48]. In previous work, we also evaluated the performance of the overall system: the ML-based analysis of encrypted data on the server required between 3.71 and 5.55 seconds, while generating the ZKP on the client took between 1.09 and 1.66 seconds [7]. These measurements were obtained using a server with 16 vCPUs and a client with 4 vCPUs. Importantly, these processing times do not interfere with usability, as authentication operates continuously in the background and decisions are made at regular intervals without disrupting user interaction.

d: GNN-Driven Predictive Authentication

As HE operations are computationally expensive, using a periodic check for the users legitimization becomes costly. One possible way to mitigate this overhead is to replace periodic checks with event-triggered authentication driven by on-device path prediction. Concretely, we predict whether the user is about to navigate into a *critical* screen (e.g., profile management or payment) and only then prompts for authentication. This concept was explored as a potential extension of the framework.

For prediction we employ GRETEL [49] a graph neural network (GNN) that extrapolates likely future navigation paths from recent interaction context. To do this, we first need to represent the Raumobil app as a graph. Let the app navigation be a directed graph $G = (V, E)$, where each node $v \in V$ is a screen and each directed edge $e_{i,j} \in E$ indicates that at least one transition from screen i to j was observed. We designate a subset $V_{\text{crit}} \subseteq V$ as critical destinations and include a special START node for session beginnings. We derive these transitions from past sessions of anonymized users who used the *regiomove* or *OrtenauMobil* app and navigated through it.

A user session is a sequence $\pi = (v_1, \ldots, v_T)$ with timestamps $(t_1, \ldots, t_T)$. For training, we take sliding prefixes $\pi_{\leq t} = (v_1, \ldots, v_t)$ and assign a binary label indicating whether a critical screen will be reached within horizon H:

$$y_t = \mathbf{1}\Big\{\exists\, \tau \in \{1, \ldots, H\} \;:\; v_{t+\tau} \in V_{\text{crit}}\Big\}.$$

The GRETEL GNN f_θ consumes G, node/edge features, and the current path prefix $\pi_{\leq t}$ to produce a probability that a critical node will be reached within H.

We train exclusively on sessions from authenticated users, since only these sessions can reach critical screens. This removes impossible transitions for anonymous sessions and reduces label noise. To improve discrimination, we augment the graph with contextual features available at inference time for example the active mobility provider, the cumulative number of prior app visits for the account, and a binary indicator of past bookings, capturing provider-specific flows, familiarity effects, and recurrence likelihood.

If the predictor does not trigger in advance, authentication is enforced deterministically upon actual entry into a critical screen. The model can only prompt earlier, never suppress mandatory checks.

D. CROSS-PROVIDER AUTHENTICATION INFRASTRUCTURE

To facilitate seamless integration of the behavioral authentication system with identity and service providers, we propose a reverse proxy named CoNym. Positioned in front of existing systems, CoNym acts as a single entry point for client applications and supports cross-provider authentication, thereby enabling integration with multiple providers. Moreover, by building on OAuth 2.0, CoNym minimizes the changes required from both clients and providers.

In standard OAuth 2.0, two authentication methods are available. First, a client can request a fresh login, prompting the user for credentials; upon success, the client receives an access token and refresh token[8]. Second, the client may use a refresh token to obtain a new access token without further user interaction. In contrast, CoNym introduces a third method: the device authenticates to CoNym, after which the behavioral data analysis results for that device produced by the process described in Section IV-C is consulted. If the behavior matches previous patterns, the user is re-authenticated, and new tokens are issued. Importantly, the two standard methods are also extended to support device authentication during login, enabling later use of behavioral authentication from that device.

Furthermore, when acting as a reverse proxy, CoNym forwards non-behavioral and non-refresh-token requests to a connected identity provider. If multiple providers are available, the client or user may select one. In this role, CoNym behaves as a standard OAuth 2.0 client towards the connected identity provider, and therefore requires no modifications to the infrastructure of these providers.

However, a key challenge arises because tokens issued by CoNym are not directly accepted by services. To address this limitation, CoNym additionally proxies service requests. Routes to existing service endpoints can be configured on CoNym to establish a proxy endpoint for them. On a request the proxy validates the CoNym-issued access tokens and exchanges it for a provider-issued token via the OAuth 2.0 Token Exchange standard [39], before it forwards the request with the exchanged token. While this requires the provider to accept CoNym-issued tokens at its Token Exchange endpoint, it allows services to operate unchanged.

Timing is critical in behavioral authentication. When CoNym receives a request, the latest analysis reflects only recently collected data, which may be insufficient immediately after application startup. Since data collection begins only at launch (to preserve battery life), early authentication attempts may fail due to a lack of data. To mitigate this, the collection window starts short—only a few seconds—and gradually increases. Although early results have lower confidence, they reduce the need for credential prompts. Moreover, because to-

[8]If OpenID Connect is enabled, an ID token is also issued.

kens are short-lived, clients re-authenticate periodically, and subsequent attempts benefit from longer collection periods and thus higher confidence.

In addition, another timing-related challenge arises from the periodic arrival of behavioral data. The most recent analysis result cannot always be reused, as it may be outdated, user behavior may have changed, or the application may have stopped transmitting data. To handle this, CoNym applies a threshold-and-decay mechanism. Each analysis produces a confidence score indicating the likelihood that the current user matches the previously authenticated one. When authentication is attempted, this score is reduced according to a decay function that accounts for the age of the result. If the decayed score remains above the threshold, authentication succeeds; otherwise, it fails. Consequently, the threshold and decay function must be carefully tuned to the data collection intervals and analysis times to ensure an optimal balance between security and usability.

E. FEDERATED LEARNING WITH DATA GOVERNANCE

As shown in Figure 1, several regional mobility providers operating in the tri-national Upper-Rhine region face the same continuous user authentication problem. Combining the data of these providers can potentially increase the quality of the Authentication AI. Direct pooling of raw data is infeasible because participating organizations are bound by legal and contractual constraints. To tackle this problem, we introduce a Hybrid Federated Learning Platform that combines central Data Governance (DG) and federated learning with decentralized Data Preprocessing. Our platform allows the providers to collaboratively train a shared authentication model without disclosing raw records.

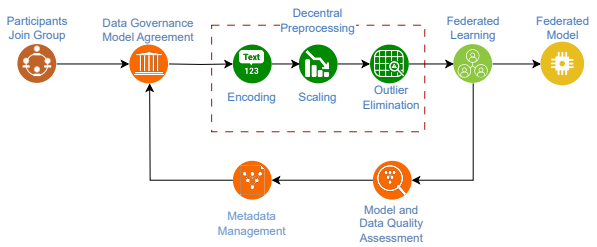


FIGURE 5: Diagram representing the lifecycle of federated learning with data governance and decentral preprocessing

In Figure 5, we show a schematic diagram of our platform. In the first step, the organizations form a federation and then work through the DG process. DG, which can be defined as the definition and implementation of methods to appropriately handle data and model assets, as well as the distribution of responsibilities among the actors involved to ensure that the methods are followed. The work of Peregrina et al. [8] adapted DG to FL to improve the coordination process needed to perform the FL training process.

This DG process is divided into three phases: A negotiation process, a contribution estimation during the FL training, and metadata management for auditing and tracking experiment results. First, the negotiation process aims to help the participants establish goals, restrictions, and define the settings of the FL training [9]. This is accomplished with the help of a web dashboard, where participants can implement the aforementioned artifacts without needing to write code.

After the negotiation phase, we perform decentralized preprocessing of training and test data across providers. Preprocessing transforms raw data into training ready features to improve comparability, stability, and convergence. In a federated setting, these must be unified across providers to avoid collisions and drifts that degrade model quality.

Our preprocessing comprises three steps. First, categorical features are encoded by mapping local string values to numeric representations. Numerical features are scaled so magnitudes are comparable during training without sharing raw statistics. Outliers are detected and, where appropriate, removed using robust statistics or isolation-based detectors to suppress noise and increase model quality.

We deliberately keep preprocessing off the central server. Centralizing this step would require the server to assemble global vocabularies and feature statistics, however this would leak direct sensitive data as for example the vocabularies contain clear text feature values. Also as shown by Vero et al. [33] giving a central entity knowledge of all of this information enables the reconstruction of training data, thus leaking local data that is stored in the database of each participant.

Instead, we adopt privacy-preserving encoders and scalers following Piotrowski et al. [14]. For categorical features, we support word embeddings via fastText[9]or PSI-based vocabulary alignment, both avoid sharing clear text values in order to create an federate wide identical encoding. For scaling, we compute z-score parameters via secure multi-party computation. Providers secret-share mean and standard deviation, thus only revealing the aggregate metrics to each provider.

After the preprocessing is completed, we run the training process based on the agreed settings during the negotiation. For this, we extend the FL training using the Flower framework [50] with functionality to run multiple processes before the training. Besides the preprocessing, we also run a script to verify that the data adheres to the schema agreed upon during the negotiation. Once everything is executed, we run multiple rounds of hyperparameter optimization using Optuna[10]. By using Optuna, we ensure the production of the best possible ML model.

Finally, during the whole process, we collect metadata in the form of provenance and experiment tracking. The former allows tracing all the operations made by the participants. We use the PROV[11] ontology to keep a trace of all operations made during the negotiation, as well as the results of the different training runs. This information can then be used by the participants for decisions during the next negotiation.

[9]https://fasttext.cc/
[10]https://optuna.org/
[11]https://www.w3.org/TR/prov-o/

We apply this federated architecture to the GNN model explained in Section IV-C. We federate across *regiomove* and *Ortenau Mobil* (cf. Section II) to increase behavioral diversity and the sample used for the training. The applications have similar navigation interfaces and users may traverse both regions, yielding complementary trajectories. Using the secure encoding preprocessing introduced earlier, each party maps its app screens to unique numerical identifiers and feature vectors locally, so no app-internal metadata or raw logs are exchanged. This cross-operator federation provides more varied interaction patterns and thus a stronger next-screen predictor for event-triggered authentication.

In this setting, we consider federations with fewer than 20 participants to be the practically relevant. First, this is a cross-silo scenario in which participants are organizations with comparatively large local data pools rather than individual small devices. Second, training is performed at the organizational level and not on mobile end-user devices, which naturally limits the number of eligible participants. Third, identifying organizations within the same domain that are both willing and technically able to collaborate in such a federation is challenging. This is particularly true in the public transportation sector, where only a limited number of operators maintain apps with sufficiently rich and comparable mobility provider coverage. Therefore, our focus is not on scalability in terms of participant count, but on secure and effective collaboration among a small set of organizations.

V. PRELIMINARY EVALUATION

Before conducting a field test of the proposed system with a real-world public transportation application, we performed a series of offline evaluations to verify the robustness of our approach and adjust the model parameters. This preliminary experiment, performed on a public benchmark dataset, served as a controlled pre-test phase to ensure that the framework was sufficiently stable and attack-resilient for an operational rollout.

A. EVALUATION METRICS

Throughout the evaluation and the field tests, we employ two common metrics to assess the performance of our behavioral authentication system: the Equal Error Rate (EER) and the Area Under the Receiver Operating Characteristic Curve (AUC) [51]. The EER corresponds to the operating point at which the false acceptance rate (FAR) and the false rejection rate (FRR) are equal, providing an indicator of overall authentication performance. Specifically, lower EER values indicate better performance. The AUC captures the overall discriminative power across all possible thresholds, with higher values indicating better separability between legitimate and impostor sessions.

B. EXPERIMENT SETUP

In this preliminary evaluation, we aimed to validate the robustness and scalability of the *aura.ai* authentication framework before its field deployment. The experiments were designed to assess the capability of the behavioral biometrics-based CA models to discriminate between genuine users and impostors under realistic operating conditions. The *BehavePassDB* dataset [10] was employed as a public benchmark for mobile behavioral biometrics. It includes background sensor signals (accelerometer, gyroscope, magnetometer, gravity) and touchscreen-based tasks such as *Keystroke*, *Readtext*, *Gallery*, and *Tap*. Data was collected from 81 participants across four sessions, each separated by at least 24 hours, enabling the evaluation of intra-user variability.

The experimental setup relied on a client–server architecture implemented in Python using TensorFlow 2.10, Scikit-learn 1.4, and NumPy 2.0 on a Debian 12 server (Intel Core i7-10700 CPU, 32 GB RAM). Two complementary models were evaluated. The first, B2CAR [12], is an LSTM-based CA framework enhanced with regularization techniques (Ridge, Lasso, ElasticNet, and Bayesian) to prevent overfitting and improve generalization under both random and skilled impostor scenarios. The second, AI-MBBCA [11], is an advanced framework combining a GA for hyperparameter optimization with an IF layer for secondary impostor detection. Both models were trained using a triplet-loss objective, optimizing the embedding space such that genuine samples cluster together while impostors are pushed apart by a GA-tuned margin. Each training session used 150 epochs, a batch size of 512, and the Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$).

The evaluation protocol followed the *BehavePassDB* standard split: 51 users for training (random impostors only), 10 for validation, and 20 for evaluation (including both random and skilled impostors). This setup served as a preparatory validation phase before deploying system in a real-world field test with live operator data.

C. EXPERIMENT RESULTS

The preliminary evaluation demonstrated the effectiveness of the AI-driven optimization and regularization strategies in enhancing model robustness for continuous authentication (CA).

In the B2CAR model [12], the integration of regularization significantly improved classification accuracy across all tasks. In the *Keystroke* task, the **ElasticNet** configuration achieved an AUC of 81.26% under random impostor conditions and 57.01% under skilled attacks, surpassing the baseline *BehavePassDB* approach by up to +15% AUC. Similarly, for the *Readtext* task, ElasticNet and Lasso achieved 80.03% and 74.84% AUC, respectively, confirming their ability to stabilize learning and mitigate overfitting. These results validate the importance of regularization in improving generalization and reducing bias toward device-specific features.

The AI-MBBCA model [11], which integrates a hybrid GA + IF design, achieved superior performance across all modalities. For the *Keystroke* task, the model reached an AUC of 93.50% for random impostors, 71.98% for skilled impostors, and 82.74% for mixed scenarios, exceeding state-of-the-art methods by +8–13% AUC. Similar improvements were

observed for the *Readtext*, *Gallery*, and *Tap* tasks, where the GA-optimized hyperparameters and IF-based anomaly scores jointly enhanced discrimination accuracy and resilience to mimicry attacks.

While the B2CAR model [12] demonstrated high stability and computational efficiency, particularly under resource-constrained conditions, the AI-MBBCA model [11] achieved the best overall performance due to its dual optimization GA and defense IF mechanisms. Together, these models validated the feasibility of deploying adaptive CA solutions within the *aura.ai* framework, ensuring strong robustness against both random and skilled impostor behaviors.

Our experiments confirmed up to a 15% improvement in AUC over baseline models, with the ElasticNet configuration achieving 81.26% AUC for random impostors and Lasso reaching 60% for skilled impostors. These results, summarized in Table 1, demonstrate that lightweight model-level enhancements can provide meaningful defense against evasion attempts, making continuous authentication more reliable in adversarial environments.

TABLE 1: AUC Performance for Accelerometer Sensor in the Keystroke Task

Approach	Random impostor	Skilled impostor
B2CAR (Ridge)	80.99%	56.55%
B2CAR (Lasso)	80.60%	**60.00%**
B2CAR (ElasticNet)	**81.26%**	57.01%
B2CAR (Bayesian)	80.90%	55.53%
BehavePassDB [10]	66.23%	56.22%

Figure 6 compares the AUC performance of **AI-MBBCA** against two reference methods [10], [13] on the *Keystroke* task using fused background sensor data. The results clearly demonstrate the effectiveness of the proposed method, achieving 93.50% AUC for random impostors, 71.98% for skilled impostors, and 82.74% for mixed cases, thus outperforming prior systems across all attack scenarios. These gains confirm that combining GA-driven optimization with IF-based anomaly detection substantially enhances the discriminative power of behavioral models, ensuring accurate and attack-resilient authentication in real-world deployments.

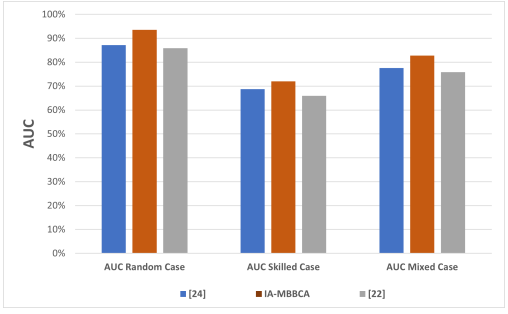


FIGURE 6: AUC performance for the best fusion of different background sensors in the Keystroke task.

VI. AUTHENTICATION FIELD TEST

To validate the entire authentication framework described in Section IV-C, we conducted a field test using the public transport mobile application *regiomove* that was described in Section II. To quantify the authentication accuracy, we rely on the metrics introduced in Section V-A.

A. APPLICATION SETUP

For our field test, we built a clone of the Android version of the *regiomove* application (see Section II). The clone used a separate user base and, although functionally equivalent, had no access to the original system or to users outside the field test. We integrated our behavioral authentication system outlined in Section IV. This was achieved by establishing an Android service that is responsible for data collection, the preprocessing pipeline and the communication with the CA Server. Additionally, we set up a monitoring of the CA Server to get notified in the case of a system outage or software errors. Further, the *regiomove* application was adjusted to establish the necessary communication with the CA service.

Furthermore, the *regiomove* application was modified so that it uses CoNym as both the identity and service provider, instead of connecting directly to the services offered (cf. Figure 2). CoNym delegates regular authentication and service requests to *INIT*, while handling behavioral authentication requests itself. The *regiomove* application was adapted to use behavioral authentication requests instead of refresh token–based ones for its authentication procedure, falling back to regular credential-based authentication only if behavioral authentication fails. Although the authentication infrastructure supports multiple identity and service providers, only *INIT* is used for the field test. In this way, we were able to set up the full behavioral authentication process of our framework (cf. Figure 2) in the context of the *regiomove* application.

B. ML MODEL TRAINING

We require ML models at two places in the framework: (1) for preprocessing behavioral data on the client device and (2) for analyzing embeddings to derive a score for authentication. As described in Section IV-C, we rely on public datasets to train the models before conducting the actual field test. Here, we used the BehavePassDB dataset [10] and the BrainRun dataset [52]. To train the models, we use the procedures introduced in Section IV-B.

C. AUTHENTICATION SETTING

In terms of the setting, we employed an initial user authentication based on the mechanism that was already part of *regiomove*, which involves sending a text message via SMS containing a one-time code to the user's device for verification. If the code is entered correctly, a token is issued to the client, which is valid for one minute. Within this phase, we collect baseline data that captures the behavior of the user and is required for the behavioral analysis (cf. Section IV-C). Subsequently, we switch to behavioral authentication using

the proposed framework. In case an authentication fails (non-legitimate behavior recognized), we fall back to the original authentication factor and send a SMS to the user device. This avoids the user being unable to log in due to potentially incorrect results from the behavioral authentication. Note that the fallback mechanism would be not suitable in a production system, as SMS verification allows an attacker with a stolen device to receive the message. We nonetheless adopted this approach in our prototype to keep the modifications to the original application minimal and to avoid negatively affecting user acceptance due to other aspects.

D. STUDY DESIGN

The field test was carried out over the duration of one month and involved 16 participants. Thirteen of them were equipped with Samsung Galaxy A16 Android devices. The remaining three participants used their own personal Android phones, although the exact device models were not tracked. The homogeneity of the devices helps minimize the impact of device-specific characteristics, ensuring that the authentication focuses on behavioral patterns rather than artifacts introduced by different hardware. During the field test period, participants regularly interacted with the application in an uncontrolled setting, completed weekly challenges, and attended a two-hour special event. Weekly challenges required each participant to complete predefined tasks in the application, such as searching for an optimal connection from one place to another or simulating behavior variations. During the special event, participants collaborated in groups and tried to access each other's phones by imitating the behavior of legitimate users. This setting is similar to the skilled attacker scenario conducted in the BehavePassDB dataset [10]. In order to be able to evaluate the accuracy of our approach in these attack scenarios, we have explicitly labeled the sessions that were part of the event. Further, during the field test, behavioral sensor data was collected and stored in unencrypted form. This was necessary because several evaluation metrics, such as AUC, as well as analyses of threshold sensitivity, require continuous authentication scores rather than only binary authentication decisions. Moreover, the recorded raw data allowed us to replay sessions and emulate the privacy-preserving authentication offline under different parameter configurations. Since prior work has shown that the HE-based analysis used in our framework has a negligible impact on authentication accuracy, this methodology does not impact the results observed during evaluation [7]. To complement these activities, interviews were conducted during the field test.

E. ETHICAL AND PRIVACY CONSIDERATIONS

Because the study includes behavioral data, suitable data protection mechanisms are required to comply with the General Data Protection Regulation (GDPR). In particular, as described in the previous section, raw behavioral sensor data was collected during the field test to enable detailed evaluation and the emulation of the privacy-preserving authentication framework. An important point for continuous authentication is that the collected data is not used to identify individuals. Instead, the data is only used to observe variations in device usage patterns to verify the legitimate user and does not allow conclusions regarding personal identity. To ensure the privacy of participants, we used pseudonyms to associate data with a specific user. Furthermore, informed consent was obtained from each participant, specifying the scope, nature, and purpose of the data collection. In addition, a Data Protection Impact Assessment (DPIA) was conducted to evaluate possible risks and define countermeasures to mitigate them.

F. FIELD TEST RESULTS

After establishing the robustness and limitations of our models through preliminary controlled evaluations, we now turn to insights gained from real-world deployment. In the following, we first summarize the behavioral data collected during the field test and then report both offline analyses, where authentication performance is evaluated retrospectively on the acquired dataset, and online results observed during live operation, where authentication decisions were made in real time using fixed system parameters.

a: Collected Data

We recorded behavioral data of 16 different devices, where each device is mapped exactly to one participant of the field test. In the first step, we removed sessions with corrupt data or unusually short durations (less than 20 seconds). As a result, 39 sessions were discarded and the number of usage sessions for each device varied between 5 and 20, with an average of 8 sessions per user. In total, 129 unique usage sessions were recorded, each lasting an average of 5.29 minutes. The special event comprised 60 sessions, 70% of which (42 sessions) simulated attack scenarios.

For each session, we collected data from multiple sensors, specifically the accelerometer, linear accelerometer, gravity sensor, gyroscope and magnetometer. For each sensor, the sampling rate was set to 100 Hz. However, depending on the specific device, some sessions lack data from certain sensors due to hardware differences or temporary sensor unavailability. Moreover, all sensor streams were timestamped and later synchronized to enable consistent feature extraction across data sources.

Overall, the collected dataset amounts to approximately 5.8 GB of raw sensor data, underlining the scale and diversity of the behavioral information gathered during the field test.

b: Offline Authentication Accuracy

In a first step, we evaluate the ability of our behavioral authentication approach to distinguish between legitimate and illegitimate users based on the data collected during the field test. This analysis is conducted offline and does not reflect the concrete authentication outcomes observed during the actual operation of the system. Instead, it serves to assess how well the proposed models and methods can separate sessions originating from the same user from those of different users

when applied to the collected dataset. To this end, we compare the similarity scores produced by the ML-based analysis for intra-user session pairs and inter-user session pairs.

In addition, we examine the impact of the decision threshold used to convert similarity scores into binary authentication outcomes. This analysis provides insight into the sensitivity of the authentication process with respect to key parameters such as the decision threshold and the weighting of individual sensor modalities, which may vary depending on the deployment environment and the characteristics of the devices used. While the field test itself was conducted with a single fixed, predefined threshold, we also determine the optimal threshold retrospectively to estimate the best possible performance that could have been achieved under ideal parameter selection. We begin by analyzing the authentication performance for each individual data source to evaluate their respective contribution to the overall process. Afterwards, we assess the overall accuracy that can be achieved with the optimal threshold.

Per-sensor performance: We first analyzed each data source individually to understand its isolated contribution to authentication. Table 2 summarizes the EER and AUC values for the five sensors. The accelerometer and magnetometer

TABLE 2: Offline EER and AUC per data source

Data source	EER [%]	AUC [%]
Accelerometer	**19.03**	**87.83**
Lin. Accelerometer	24.89	83.67
Gravity Sensor	35.24	71.52
Gyroscope	46.18	56.74
Magnetometer	19.19	87.60

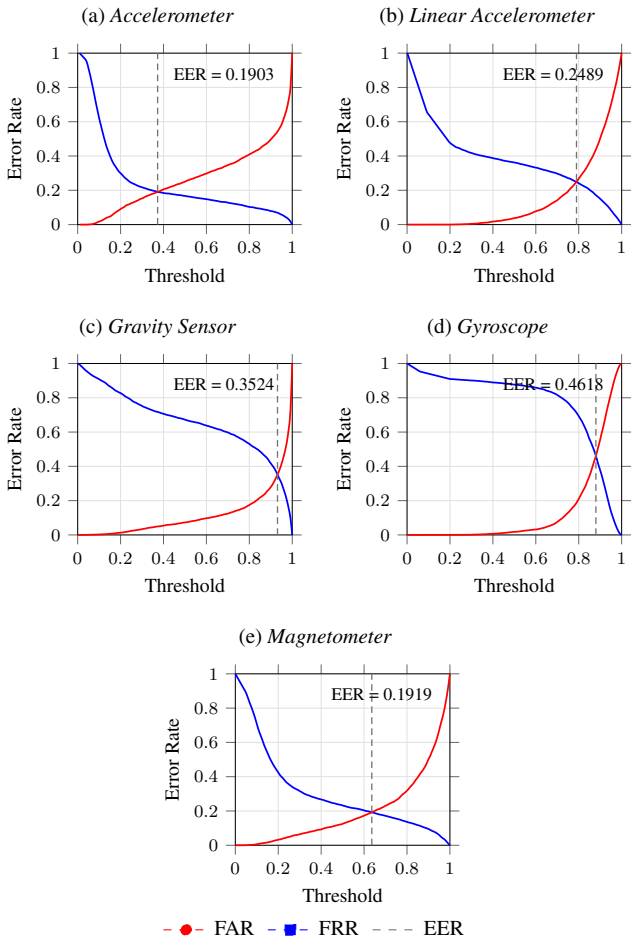


FIGURE 7: Comparison of the authentication performance reached by different data sources on the behavioral data collected in the field test

data achieve the best discriminative performance, with EERs of 19.03% and 19.19% and AUC values of 87.83% and 87.60%, respectively. In contrast, the linear accelerometer and the gravity sensor perform slightly worse, with EERs between 24.89% and 35.24%. It also turns out that, in our setting, the gyroscope does not reflect any user behavior characteristics that are valuable for authentication, with an EER and AUC of close to 50%. Overall, these results indicate that accelerometer and magnetometer are particularly informative for user differentation in the public transport application used.

Since the EER and AUC only indicate the quality of the authentication performance independently of the threshold, we visualized the false acceptance rate (FAR) and false rejection rate (FRR) for all data sources in Figure 7. It shows the FAR and FRR values (y-axis) depending on the selected threshold (x-axis) to estimate how many legitimate users are rejected and how many illegitimate users are accepted for a certain threshold. Note that the intersection of the FAR and FRR here represents the EER. The figure indicates how important it is to set the threshold properly, because if it is set too high, for example, legitimate users will be correctly recognized, but a lot of attackers can bypass the authentication. This is a scenario needs to be avoided in real-world systems. It can be seen that the threshold for achieving the EER varies greatly between the different data sources. For example, the accelerometer data reaches a low FRR already at a significantly lower threshold than the magnetometer, despite achieving similar performance overall. These results again highlight the importance of the choice of the threshold. Further, this task gets even more complex, when considering the combination of different data sources.

In the final step, we assessed the accuracy of detecting the simulated attack scenarios across the various data sources. Here, we analyzed the accuracy for different thresholds: (1) ACC@EER, representing accuracy at the equal error rate; (2) ACC@FPR10, and (3) ACC@FPR30, corresponding to fixed false positive rates of 10%, and 30%, respectively. The results are shown in Table 3.

The results show that both the accelerometer and magnetometer achieve the highest accuracy across all thresholds, confirming their strong discriminative power for behavioral authentication. As expected, accuracy increases with more permissive false positive rates (FPR), with ACC@FPR30 reaching over 90% for these two sensors. In contrast, the gyroscope and gravity sensor exhibit limited discriminative ca-

TABLE 3: Accuracy for detecting imitation attacks at different thresholds per data source

Data source	ACC@EER	ACC@FPR10	ACC@FPR30
Accelerometer	81.09	54.84%	92.48%
Lin. Accelerometer	75.50%	33.04%	87.22%
Gravity Sensor	65.71%	30.64%	60.68%
Gyroscope	53.91%	11.24%	35.30%
Magnetometer	79.96%	50.98%	92.42%

pability, indicating that motion and orientation data alone are insufficient for reliable authentication in this context. Moreover, the results demonstrate that the proposed behavioral authentication approach can accurately detect even strong attackers who attempt to imitate legitimate users, emphasizing its robustness against real-world threats.

Overall performance: We then combined all data sources to compute a single authentication score per session. In doing so, we assumed the optimal threshold ($othr$) and also an optimal weighting $(w_1, \ldots, w_5)$ of the various data sources (not a simple average across all data sources). More specifically, the authentication score was calculated as $as = w_1 * s_{acc} + w_2 * s_{accl} + w_3 * s_{grav} + w_4 * s_{gyro} + w_5 * s_{magn}$, where $s_{acc}, s_{accl}, s_{grav}, s_{gyro}, s_{magn}$ denote the scores for the accelerometer, linear accelerometer, gravity sensor, gyroscope and magnetometer, respectively. The optimal weights determined were: $(w_1, w_2, \ldots, w_5) = (+2.885114, +2.633770, +0.0, +0.0, +3.071426)$. Consequently, both the gyroscope and the gravity sensor did not contribute to the final authentication score, indicating that their inclusion did not improve overall performance during optimization.

After determining the optimal weights for combining the individual data sources, we evaluated the overall performance of the fused authentication model. The resulting weighted combination achieved an **AUC of 91.81%** and an **EER of 15.03%** at the optimal threshold of approximately 0.63, clearly outperforming any individual sensor. At this threshold, the system reached an **overall accuracy of 84.97%**, confirming that the fusion of multiple complementary signals leads to a more robust authentication decision. The attack detection performance further supports this conclusion, with a **true positive rate (TPR)** of **0.7225** at a **10% false positive rate (FPR)**, **0.9142** at a **20% FPR**, and **0.9729** at a **30% FPR**. These results demonstrate that the fusion can substantially improve discriminative power and resilience against impersonation attempts. In summary, the combination of accelerometer and magnetometer signals provides strong authentication performance in our setting and confirms the viability of multimodal sensor fusion in real-world mobile applications for public transport.

c: Online Authentication Accuracy

During the execution of the field test, we had to define a fixed threshold that is used to determine whether a legitimate user is present. Based on previous results obtained on public datasets, we decided to set it to $thr = 0.8$. As described in Section IV, the CA Server sends the average authentication score calculated across the different data sources to Conym. Subsequently, Conym adds a certain decay, based on the time interval between receiving the score and the current authentication process (cf. Section IV-D).

In practice, however, the authentication performance observed during the field test deviated notably from the offline results. Specifically, the TNR, i.e., the proportion of legitimate users correctly recognized, reached 82.19%, while the TPR, i.e., the proportion of attackers correctly rejected, was considerably lower at 19.08%. These deviations from the optimal results observed during offline evaluation demonstrate the system's sensitivity to parameter choices, including decision thresholds, decay functions, and the weighting of sensor modalities. In practical deployments, careful calibration of these parameters is necessary to account for differences in user behavior, device characteristics, and environmental conditions. Overall, the online accuracy amounted to 78.96%, which was clearly below the estimated optimal accuracy. The limited attacker detection capability became particularly evident during the special event conducted at the end of the field test, in which users intentionally attempted to imitate others' behavior. Several participants were able to bypass authentication in this scenario, which was also noticed by legitimate users and negatively affected their perception of security, as later reflected in the post-study survey.

The analysis of these results revealed two primary causes for the reduced online accuracy. First, all data sources were weighted equally in the live system, although the offline analysis showed that the gravity sensor and gyroscope provided little to no discriminative information and even degraded performance when included. Second, the fixed threshold of 0.8 was set too high for the most informative data sources (accelerometer, linear accelerometer, and magnetometer), which resulted in low FRRs for legitimate users but high FARs for attackers. Consequently, while legitimate users were authenticated smoothly, attackers could occasionally be accepted as well, explaining the imbalance between TPR and TNR observed in the field. These findings are fully consistent with the offline evaluation results, which already indicated that optimal performance could only be achieved with adjusted thresholds and selective weighting of the most discriminative sensors. Moreover, the decay mechanisms increases scores as the time since the last update grows; it introduces additional temporal variance in the effective comparison values, which complicates parameterization and makes selecting a single appropriate threshold even more challenging under real-time conditions. The key insights derived from these observations are further discussed in the last paragraph of this section, which summarizes practical recommendations to help future studies avoid similar challenges in real-world behavioral authentication deployments.

d: Authentication Efficiency

We also evaluated the efficiency of our framework by measuring the execution times of the different parts of the authentica-

tion process on both a client device and a virtual CA server. The CA Server was deployed on a virtual machine with 16 vCPUs and 32 GB of random access memory. The execution times of the client tasks were recorded on a Samsung Galaxy A16 device.

For this evaluation, we excluded the network communication times and the components related to Conym. A detailed analysis of the network traffic is provided in the original paper that introduced the authentication protocol [7], while Conym is based on an efficient OAuth 2.0 implementation[12] and does not contribute noticeably to the total latency. That work also discusses the influence of parameters such as the number of data sources and configuration choices on the overall authentication latency. The results in Table 4 show that

TABLE 4: Measured execution times in seconds for the different parts of the authentication process of our framework

Entity	Task	*min*	μ	*max*	σ
Client	Preprocessing	0.02	0.02	0.1	<0.01
Client	Encryption	0.03	0.04	0.08	<0.01
Client	ZK Proof	1.22	1.46	1.74	1.11
Server	HE ML Analysis	4.01	4.57	5.37	0.24
Server	ZK Verify	<0.01	<0.01	0.01	< 0.01
Overall		**5.33**	**6.1**	**7.17**	**1.14**

the most time-consuming components of the authentication process are the generation of the ZK proof on the client side and the HE-based machine learning analysis on the CA server. In particular, the client-side ZK proof computation takes on average about 1.46 seconds, while the HE ML analysis on the server requires approximately 4.57 seconds. The other tasks, such as preprocessing, encryption, and ZK verification, contribute only marginally to the overall execution time.

It is important to note that these operations are performed in the background to ensure a smooth token renewal (see Section IV). Hence, their latency does not directly affect the user experience, since the authentication update can be triggered early enough to ensure uninterrupted access.

Furthermore, the overall latency could be significantly reduced by using more powerful server hardware. In particular, the HE ML analysis is highly parallelizable, and its runtime can decrease substantially with increased computational resources [7].

Overall, the results indicate that the authentication process is computationally feasible on standard hardware, confirming the suitability of our approach for privacy-preserving authentication with behavioral biometrics.

e: Authentication Infrastructure

In total, 760 authentication requests were processed using behavioral authentication. Because access tokens had a lifetime of 60 seconds, many of these requests were re-authentications triggered by user actions after the access token expired rather than initial logins. Of the 760 requests, 441 failed and 319 succeeded. Among the failures 284 occurred because no behavioral data was yet available for the requesting device. This is expected during initial app startup, when data collection is still incomplete (see Section IV-D), the issue is generally resolved once sufficient data accumulate.

Excluding these data-availability cases, 319 authentications exceeded the decision threshold, while 157 fell below it. This distribution suggests that the threshold and decay parameters require further calibration to improve the true acceptance rate.

Additionally, 123 authentications were delegated to the backing identity provider, of which 120 succeeded. This demonstrates that the fallback mechanism operated reliably when needed for initial logins or when the behavioral authentication was not yet feasible.

The exchange of access tokens and forwarding of requests to the backend services functioned reliably. Of the 3995 requests processed, only 91 failed. All failures were warranted and caused by client requests that tried to make a request with a invalid access token after a previous behavioral authentication failed or the token was expired. Further investigation is required to determine the underlying cause of these invalid access attempts.

f: Survey

To collect feedback from the participants, eight semi-structured interviews were conducted. Interviews showed that participants perceived the continuous authentication positive and innovative. Recommendations for improvements are unreliability of the system and missing transparency of decision making processes. 75 % of the participants reported that the system not always recognized an attacker, which led to a reduction in the perceived level of protection. As a result, one user reported a shift in the behavior by adapting interactions with the smartphone, e.g., the smartphone was kept steady to minimize motion sensor input. We consider the reported technical challenges as a results of the prototype application used in the field test and anticipate that they can be resolved in real-world applications. However, interviews suggest that continuous authentication is generally well received and can enhance users' sense of security when unauthorized individuals are detected. At the same time, users are sensitive to errors, both for false positives, i.e., when unauthorized users are not recognized and false negatives, i.e., when they themselves are logged out.

G. LESSONS LEARNED

Within the process of running our field test we noticed several key points that either were crucial to ensure that no critical failures occurred or had a major impact on the outcome of the study. In the following we will list and describe these takeaways to guide future experiments in the field of behavioral authentication:

- **Pre-field test:** In order to rule out potential sources of error, it makes sense to roll out the system internally to a small group of users before the planned field test (*pre-field test*). This allows errors in various components to be eliminated at an early stage, ensuring that the field

[12]https://docs.jans.io/

test runs as planned. In particular, the legal basis (data protection) should be explored and prepared in advance, as this can take some time, and the necessary effort should not be underestimated.

- **Fixed thresholds:** It naturally can happen that even if the ML models used generalize well to unseen users, the thresholds for making an accurate authentication decision can vary significantly. This can lead to either a lot of rejections of legitimate users or the acceptance of illegitimate users (cf. Section VI-F). For this reason, we propose to use techniques to dynamically adjust the thresholds [53], [54], conduct a pre-field test as mentioned before, or monitor the outcomes of the authentication procedure continuously to be able to adjust the thresholds if necessary. More generally, our results indicate that the robustness of behavioral authentication systems depends strongly on the calibration of several system parameters, including decision thresholds, score decay mechanisms, and the weighting of sensor modalities, which need to be adapted to the characteristics of the deployment environment.
- **Fallback authentication methods:** behavioral authentication will offer a viable alternative to traditional authentication methods in the future. Nevertheless, depending on the setting, incorrect authentication decisions may still occur. For this reason, an alternative login method should be offered, especially in cases where a legitimate user is rejected. Otherwise, this can lead to frustration among users, which weakens the acceptance of behavioral authentication methods.
- **System monitoring:** Monitoring the various components used in field testing is essential in order to be able to intervene quickly in the event of unexpected errors. In our field test, we had a case where an older package version on the CA server caused errors with some clients and ultimately lead to a crash of the CA server. With the help of monitoring, we were able to fix the error within 15 minutes and minimize the impact on the outcome of the field test.
- **Data privacy:** Data protection is a very important issue, which was actively questioned by users and had a positive influence on the acceptance of the system (see surveys in Section VI-F). This highlights the importance of privacy-protecting authentication systems that support behavioral authentication.
- **Cross-device testing:** In preparation for our field test, we conducted cross-device testing and noticed notable differences between devices, for example in sensor characteristics, sampling frequencies, and hardware precision. While these variations did not had a major impact on authentication accuracy, they represent important factors that need to be considered during data preprocessing and model development. Since our field test was carried out primarily on a homogeneous set of devices, we recommend that experiments explicitly include a diverse range of devices to ensure robustness and to identify potential device-specific effects early on.
- **Transparency and user trust:** During the field test, several participants expressed uncertainty about how and when authentication decisions were made. While continuous authentication operates in the background to minimize user disruption, a lack of transparency can reduce user trust when unexpected logouts occur. Future systems should therefore provide clearer feedback mechanisms, for example by notifying users when behavioral verification fails or when additional authentication is triggered. Such explanations can help users better understand system behavior and improve trust in continuous authentication mechanisms.

VII. FEDERATED MACHINE LEARNING FIELD TEST

To evaluate how we improve the applicability of FL to real scenarios, we run an evaluation of the Federated Machine Learning Data Governance (FML-DG) platform. We do so by performing a second field study at the University of Cádiz, where we collaborate with European partners to evaluate the following: the usability of the governance platform, how helpful it is to set up a common dataset, and how quickly FL partners can run a first training process. Therefore, we design the following experiment for the second field test. We request a group of students to make groups of three, and provide each student in the group with a partition of a dataset. This dataset, which is an actual partition of a dataset, has then been modified artificially, by both adding some columns and sorting the columns in random order. The students were then requested to communicate with each other via text messages, and then use either the platform or a notebook and Flower code that only needed some configuration steps to run. Each group would then need to agree on a common dataset structure and minimum statistics to preprocess the dataset. Then, they would need to run the training. Two sessions were run, each with 5 groups, two using the notebook and Flower code, and three using the platform. Before the experiments, the student filled out a questionnaire, where only half had some experience with Python, and only 2 out of 30 had some experience with ML. None of them had worked with FL before, and only one knew about the term. In total, half of the groups managed to finish the experiment by running the training, distributed in half the groups for the platform, and half the groups for the Flower. During the experiment, we measured how much time it took for the groups to put together a common dataset, and we set NA for those who did not manage to do so. They can be seen in Table 5. It should also be emphasized that the Flower group received an already working code for running the training. In normal conditions, they should also have to develop the code, while the platform already provides all functionality.

After the experiment, we ran two new questionnaires: one for the platform users, and another one for the Flower ones. Overall, we got mixed opinions on the usability of the platform, as not many students had prior experience with ML. Still, the majority of them found the platform usable, and

Platform	Ours						Flower			
Group	1	2	3	4	5	6	1	2	3	4
Time (in minutes)	124	96	-	-	-	92	-	122	-	89

TABLE 5: Table with the times that each group took to put together the dataset. Empty means they did not achieve to do so in the time of the experiment.

considered that the main problem using it was the complexity of the problem present already in ML. The opinions that we got from the Flower point in this direction as well. In the future, we will work on extending the experiments to consider quality improvements, as well as hyperparameter search.

VIII. CONCLUSION

In this paper, we presented a privacy-aware behavioral authentication framework developed in the context of the *aura.ai* research project [13]. The framework integrates privacy-preserving authentication, federated learning, robustness optimization, and interoperable infrastructure into a unified architecture for public transport applications. Behavioral data collected on user devices is preprocessed locally and encrypted using homomorphic encryption before transmission, ensuring that sensitive information remains confidential during analysis. The encrypted data is then processed by machine learning models on the server side to verify behavioral consistency with legitimate users. To maintain robustness against random and skilled impersonation, the framework employs optimized neural architectures combined with regularization and anomaly detection techniques. In parallel, federated learning with coordinated data governance enables multiple mobility providers to collaboratively train shared models without exchanging raw data. Finally, an authentication proxy infrastructure extends established OAuth 2.0 and OpenID Connect protocols, enabling cross-provider interoperability and seamless integration into existing mobility ecosystems. Together, these components form a deployable end-to-end architecture that demonstrates how privacy-preserving behavioral authentication can be integrated with existing authentication infrastructures while supporting collaborative model training across multiple providers.

To evaluate the practical feasibility of the proposed framework, we conducted a comprehensive evaluation consisting of a preliminary analysis, a real-world field test using the *regiomove* public transport application[14] and a second field study assessing the feasibility of the federated learning setup. The results demonstrate that our ML-based analysis achieves strong discriminative performance, while maintaining reasonable computational overhead. The fusion of accelerometer and magnetometer data yielded an AUC of 91.81% and an EER of 15.03%, confirming the benefits of multimodal behavioral analysis. Furthermore, our robustness strategies, including regularization and anomaly detection, improved resistance against both random and skilled impersonation attacks. The authentication infrastructure achieved reliable operation with acceptable latencies, demonstrating the feasibility of privacy-preserving behavioral authentication on commodity hardware. The federated learning field test further illustrated the practical feasibility of the proposed data governance approach, demonstrating that participants with limited prior experience in federated learning were able to coordinate dataset preparation and initiate a distributed training process within a short time frame. User feedback confirmed that continuous authentication is generally well received, although threshold calibration and transparency of decisions remain areas for refinement. These observations provide insights into the deployment of behavioral authentication in real-world mobility applications and highlight the importance of system-level integration across authentication infrastructure, behavioral analysis, and collaborative model training.

Future research could explore adaptive thresholding mechanisms and online learning approaches to dynamically adjust model parameters and thresholds based on evolving behavioral and contextual patterns. In addition, improving transparency of authentication decisions and providing clearer explanations of system behavior may help strengthen user trust in continuous behavioral authentication systems. Additional studies could investigate scalability in larger federated learning networks and broader mobility ecosystems involving multiple service providers. Advances in homomorphic encryption and lightweight neural network architectures may also reduce computational overhead and latency, paving the way for broader deployment of privacy-preserving behavioral authentication in mobility, internet of things, and other distributed application domains.

USE OF GENERATIVE AI

The authors used ChatGPT4 and ChatGPT5 in Section I, Section III, Section IV, Section VI, Section VII, Section VIII, and the abstract to rephrase awkward wording, improve sentence structure and shorten texts.

ACKNOWLEDGMENT

This work was supported by the aura.ai research project co-funded by the European Union through the Interreg Upper Rhine program. The authors would like to thank KVV[15], for the possibility to test the continuous authentication in the regiomove application and their support in the field test. Ortenaukreis[16] for the opportunity to use their data from OrtenauMobil, Raumobil GmbH[17] for their collaboration in integrating and testing the behavioral authentication framework within the regiomove public transport application, and INIT[18] for connecting the authentication mechanisms with the underlying infrastructure of *regiomove*. The authors also thank Neomia.ai[19] for their valuable guidance in the design

[13] https://www.h-ka.de/en/iaf/aura-ai
[14] https://www.kvv.de/mobilitaet/regiomove.html
[15] https://www.kvv.de/
[16] https://www.ortenaukreis.de/
[17] https://www.raumobil.com/
[18] https://www.initse.com/ende/home/
[19] https://www.neomia.ai/en/

of the machine learning model architectures and their contributions to the review and refinement of the overall system concept.

REFERENCES

[1] M. Abuhamad, A. Abusnaina, D. Nyang, and D. Mohaisen, "Sensor-based continuous authentication of smartphones' users using behavioral biometrics: A contemporary survey," *IEEE Internet of Things Journal*, vol. 8, no. 1, pp. 65–84, 2021.

[2] A. Berson, *Client-server architecture*, 2nd ed. McGraw-Hill, 1992.

[3] L. Hernández-Álvarez, J. M. de Fuentes, L. González-Manzano, and L. Hernández Encinas, "Privacy-preserving sensor-based continuous authentication and user profiling: A review," *MDPI Sensors*, vol. 21, no. 1, p. 92, 2021.

[4] D. Progonov, V. Cherniakova, P. Kolesnichenko, and A. Oliynyk, "Behavior-based user authentication on mobile devices in various usage contexts," *Springer EURASIP Journal on Information Security*, vol. 2022, no. 1, p. 6, 2022.

[5] D. Monschein, J. A. Peregrina Pérez, T. Piotrowski, Z. Nochta, O. P. Waldhorst, and C. Zirpins, "Towards a peer-to-peer federated machine learning environment for continuous authentication," in *Proc. IEEE Symp. on Computers and Communications (ISCC)*, Virtual Event, 2021, pp. 1–6.

[6] W. Yang, S. Wang, H. Cui, Z. Tang, and Y. Li, "A review of homomorphic encryption for privacy-preserving biometrics," *MDPI Sensors*, vol. 23, no. 7, p. 3566, 2023.

[7] D. Monschein, A. Niedermayer, and O. P. Waldhorst, "PPMLAuth: Privacy-preserving and tamper-resistant behavioral authentication through machine-learning–based data fusion," 2025, manuscript under review. [Online]. Available: https://osf.io/h6vab/?view_only=0bb64e5cb2fc4942a86babd9e98d8980

[8] J. A. Peregrina, G. Ortiz, and C. Zirpins, "Towards data governance for federated machine learning," in *Proc. European Conf. on Service-Oriented and Cloud Computing (ESOCC)*, 2022, pp. 59–71.

[9] ——, "A Platform to Integrate Data Governance in Federated Learning." Dubrovnik, Croatia: IEEE, Oct. 2025, (Accepted).

[10] G. Stragapede, R. Vera-Rodriguez, R. Tolosana, and A. Morales, "BehavePassDB: Public database for mobile behavioral biometrics and benchmark evaluation," *Elsevier Pattern Recognition*, vol. 134, p. 109089, 2023.

[11] M. Al Samara, I. Bennis, M. Gilg, B. Brik, and A. Abouaissa, "Ai-driven optimisation for mobile behavioural biometrics continuous authentication," in *2025 International Conference on Wireless and Mobile Computing, Networking and Communications. (WiMob)*. IEEE (Accepted), 2025.

[12] M. Al Samara, M. Gilg, A. Abouaissa, I. Bennis, and P. Lorenz, "B2CAR: Behavioural biometrics for continuous authentication with regularisation techniques," in *Proc. 21st IEEE Int. Conf. on Wireless Communications and Mobile Computing (IWCMC)*, 2025, pp. 324–329.

[13] D. Monschein and O. P. Waldhorst, "Optimizing privacy-preserving continuous authentication of mobile devices," in *Proc. 18th Int. Conf. on Network and System Security (NSS)*, 2024, pp. 63–81.

[14] T. Piotrowski, Z. Nochta, M. Karl, and M. Johns, "Privacy-preserving encoding and scaling of tabular data in horizontal federated learning systems," in *Proc. Int. Conf. on Availability, Reliability and Security (ARES)*, 2025, pp. 402–424.

[15] D. Deb, A. Ross, A. K. Jain, K. Prakah-Asante, and K. V. Prasad, "Actions speak louder than (pass) words: Passive authentication of smartphone users via deep temporal features," in *Proc. IEEE Int. Conf. on Biometrics (ICB)*, 2019, pp. 1–8.

[16] M. Abuhamad, T. Abuhmed, D. Mohaisen, and D. Nyang, "AUToSen: Deep-learning-based implicit continuous authentication using smartphone sensors," *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 5008–5020, 2020.

[17] A. Acien, A. Morales, R. Vera-Rodriguez, and J. Fierrez, "Smartphone sensors for modeling human-computer interaction: General outlook and research datasets for user authentication," in *Proc. 44th IEEE Annual Computers, Software, and Applications Conf. (COMPSAC)*, 2020, pp. 1273–1278.

[18] K. S. Balagani, P. Gasti, A. Elliott, A. Richardson, and M. O'Neal, "The impact of application context on privacy and performance of keystroke authentication systems," *IOS Press Journal of Computer Security*, vol. 26, no. 4, pp. 543–556, 2018.

[19] A. F. Baig, S. Eskeland, and B. Yang, "Novel and efficient privacy-preserving continuous authentication," *MDPI Cryptography*, vol. 8, no. 1, pp. 1–14, 2024.

[20] S. F. Shahandashti, R. Safavi-Naini, and N. A. Safa, "Reconciling user privacy and implicit authentication for mobile devices," *Elsevier Computers & Security*, vol. 53, pp. 215–233, 2015.

[21] B. Topcu, C. Karabat, M. Azadmanesh, and H. Erdogan, "Practical security and privacy attacks against biometric hashing using sparse recovery," *Springer EURASIP Journal on Advances in Signal Processing*, vol. 2016, no. 1, p. 100, 2016.

[22] P. Lacharme, E. Cherrier, and C. Rosenberger, "Preimage attack on Bio-Hashing," in *Proc. Int. Conf. on Security and Cryptography (SECRYPT)*, 2013, pp. 1–8.

[23] G. Vassallo, T. Van Hamme, D. Preuveneers, and W. Joosen, "Privacy-preserving behavioral authentication on smartphones," in *Proc. 1st ACM Int. Workshop on Human-Centered Sensing, Networking, and Systems (HumanSys)*, 2017, pp. 1–6.

[24] M. Bellare, T. Ristenpart, P. Rogaway, and T. Stegers, "Format-preserving encryption," in *Proc. 16th Int. Workshop on Selected Areas in Cryptography (SAC)*, 2009, pp. 295–312.

[25] J. Domingo-Ferrer, Q. Wu, and A. Blanco-Justicia, "Flexible and robust privacy-preserving implicit authentication," in *Proc. 30th IFIP Int. Conf. on ICT Systems Security and Privacy Protection (SEC)*, 2015, pp. 18–34.

[26] A. S. Beggari, A. Wali, A. Khaldi, M. R. Kafi, and A. K. Sahu, "Secure and imperceptible medical image watermarking via multiscale qr embedding and attention-based optimization," *Elsevier Engineering Science and Technology, an International Journal*, vol. 73, p. 102250, 2026.

[27] ——, "Fdct-based watermarking for robust and imperceptible medical image protection," *Elsevier Intelligence-Based Medicine*, vol. 12, p. 100280, 2025.

[28] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-efficient learning of deep networks from decentralized data," 2017. [Online]. Available: https://arxiv.org/abs/1602.05629

[29] M. Schlegel, D. Scheliga, K.-U. Sattler, M. Seeland, and P. Mäder, "Collaboration management for federated learning," in *Proc. 40th IEEE Int. Conf. on Data Engineering Workshops (ICDEW)*, 2024, pp. 291–300.

[30] B. A. Galende, S. U. Mayoral, F. M. García, and S. B. Lottmann, "FLIP: A new approach for easing the use of federated learning," *MDPI Applied Sciences*, vol. 13, no. 6, p. 3446, 2023.

[31] A. Bakarsky, D. I. Dimitrov, M. Baader, and M. Vechev, "SPEAR++: Scaling gradient inversion via sparsely-used dictionary learning," 2025. [Online]. Available: https://arxiv.org/abs/2510.24200

[32] L. H. Fowl, J. Geiping, W. Czaja, M. Goldblum, and T. Goldstein, "Robbing the Fed: Directly obtaining private data in federated learning with modified models," in *Proc. Int. Conf. on Learning Representations (ICLR)*, 2022.

[33] M. Vero, M. Balunović, D. I. Dimitrov, and M. Vechev, "TabLeak: Tabular data leakage in federated learning," in *Proc. 40th Int. Conf. on Machine Learning (ICML)*, 2023, pp. 31 069–31 101.

[34] T. Wink and Z. Nochta, "An approach for peer-to-peer federated learning," in *Proc. IEEE/IFIP Int. Conf. on Dependable Systems and Networks Workshops (DSN-W)*, 2021, pp. 150–157.

[35] D. Pasquini, M. Raynal, and C. Troncoso, "On the (in)security of peer-to-peer decentralized machine learning," in *Proc. IEEE Symp. on Security and Privacy (S&P)*, 2023, pp. 418–436.

[36] D. Hardt, "The oauth 2.0 authorization framework (rfc 6749)," 2012. [Online]. Available: https://www.rfc-editor.org/info/rfc6749

[37] N. Sakimura, J. Bradley, M. Jones, B. de Medeiros, and C. Mortimore, "Openid connect core 1.0 final specification," 2023. [Online]. Available: https://openid.net/specs/openid-connect-core-1_0.html

[38] A. Parecki, G. Fletcher, and P. Kasselman, "Oauth 2.0 for first-party applications," 2025. [Online]. Available: https://datatracker.ietf.org/doc/draft-ietf-oauth-first-party-apps/01/

[39] M. B. Jones, A. Nadalin, B. Campbell, J. Bradley, and C. Mortimore, "Oauth 2.0 token exchange," 2020. [Online]. Available: https://www.rfc-editor.org/info/rfc8693

[40] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song, "Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 1, pp. 136–148, 2013.

[41] A. Acien, A. Morales, R. Vera-Rodriguez, J. Fierrez, and R. Tolosana, "Multilock: Mobile active authentication based on multiple biometric and behavioral patterns," in *1st International Workshop on Multimodal*

Understanding and Learning for Embodied Applications. Association for Computing Machinery, 2019, p. 53–59.

[42] T. Mustafić, A. Messerman, S. A. Camtepe, A.-D. Schmidt, and S. Albayrak, "Behavioral biometrics for persistent single sign-on," in *Proceedings of the 7th ACM Workshop on Digital Identity Management*, 2011, p. 73–82.

[43] S. J. Preece, J. Y. Goulermas, L. P. J. Kenney, and D. Howard, "A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data," *IEEE Trans. on Biomedical Engineering*, vol. 56, no. 3, pp. 871–879, 2009.

[44] J. H. Cheon, A. Kim, M. Kim, and Y. Song, "Homomorphic encryption for arithmetic of approximate numbers," in *Proc. 23rd IACR Int. Conf. on Theory and Application of Cryptology and Information Security (ASIACRYPT)*, 2017, pp. 409–437.

[45] C. P. Sah, M. Kaur, and G. Singh, "Efficiency of zero-knowledge proofs: A thorough review and analysis," in *Proc. 5th IEEE Int. Conf. on Public Key Infrastructure and its Applications (PKIA)*, 2024, pp. 1–7.

[46] B. Oude Roelink, M. El-Hajj, and D. Sarmah, "Systematic review: Comparing zk-SNARK, zk-STARK, and bulletproof protocols for privacy-preserving authentication," *Wiley Security and Privacy*, vol. 7, no. 5, p. e401, 2024.

[47] J. Yuan, W. Liu, J. Shi, and Q. Li, "Approximate homomorphic encryption based privacy-preserving machine learning: A survey," *Springer Artificial Intelligence Review*, vol. 58, no. 3, p. 82, 2025.

[48] C. Lin, J. He, C. Shen, Q. Li, and Q. Wang, "CrossBehaAuth: Cross-scenario behavioral biometrics authentication using keystroke dynamics," *IEEE Trans. on Dependable and Secure Computing*, vol. 20, no. 3, pp. 2314–2327, 2023.

[49] J.-B. Cordonnier and A. Loukas, "Extrapolating paths with graph neural networks," 2019. [Online]. Available: https://arxiv.org/abs/1903.07518

[50] D. J. Beutel, T. Topal, A. Mathur, X. Qiu, J. Fernandez-Marques, Y. Gao, and et al., "Flower: A friendly federated learning research framework," 2022. [Online]. Available: https://arxiv.org/abs/2007.14390

[51] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Springer Nature Scientific Reports*, vol. 14, no. 1, p. 6086, 2024.

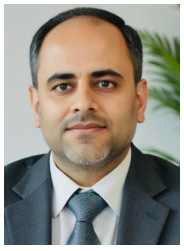
[52] M. D. Papamichail, K. C. Chatzidimitriou, T. Karanikiotis, N.-C. I. Oikonomou, A. L. Symeonidis, and S. K. Saripalle, "BrainRun: A behavioral biometrics dataset towards continuous implicit authentication," *MDPI Data*, vol. 4, no. 2, pp. 1–17, 2019.

[53] M. Smith-Creasey and M. Rajarajan, "Adaptive threshold scheme for touchscreen gesture continuous authentication using sensor trust," in *Proc. 16th IEEE Int. Conf. on Trust, Security and Privacy in Computing and Communications (TrustCom)*, 2017, pp. 554–561.

[54] W. Verheyen, "Adaptive thresholding for fair and robust biometric authentication," in *Proc. 24th Int. Middleware Conf.: Demos, Posters and Doctoral Symp. (Middleware)*, 2023, pp. 7–8.

DAVID MONSCHEIN received his B.Sc. and M.Sc. in Computer Science from the Karlsruhe Institute of Technology (KIT), Germany, in March 2018 and December 2020, respectively. He is currently pursuing his Ph.D. with the Institute of Data-Centric Software Systems (IDSS) at the Karlsruhe University of Applied Sciences, Germany. His research focuses on privacy-preserving machine learning techniques for user authentication. Besides, his interests include machine learning in general, performance modeling, and homomorphic encryption schemes.

MARKUS KNECHT received his B.Sc. and M.Eng. in Computer Science from the University of Applied Sciences Northwestern Switzerland (FHNW), and a second M.Sc. and Ph.D. in Computer Science from the University of Zürich (UZH). He completed his Ph.D. in February 2025, focusing on the design and development of a secure smart contract programming to develope robus blockchain applications. He works at the FHNW Institute for Mobile and Distributed Systems (IMVS). His research interests include blockchains, programming language & security protocol design, cryptography, and zero-knowledge proofs.



MUSTAFA AL SAMARA received his bachelor's degree in information technology engineering from Damascus University, Syria, in 2006. Between 2007 and 2010, he served as a System and Network Administrator at the Commercial Bank of Syria. Following that, from 2010 to 2013, he worked as a lecturer at the Arab International University (AIU) in Syria. He earned a master's degree in information technology engineering from Damascus University, focusing on VoIP Security, in 2011. Furthermore, he holds a Diploma in Methods Informatics Applied to Management from the University of Paris 1, Pantheon Sorbonne, obtained in 2014. Between 2020 and 2023, he earned his Ph.D. in Information Technology from the University of Haute-Alsace (UHA), France. Between September 2023 and Augest 2024, he worked as a temporary professor for research and teaching (A.T.E.R) at the University of Strasbourg. Since September 2024, he has been working as Postdoctoral Researcher at the University of Haute Alsace (UHA). His research interests include anomaly detection and Federated learning issues in WSN and IoT networks. Dr. Al Samara has authored several articles published in peer-reviewed international journals and conferences.

JOSE A. PEREGRINA José Antonio Peregrina was born in Cádiz, 1996. He is a Ph.D. student at the University of Cádiz and a research assistant member of the Institute of Data-Centric Software Systems at Karlsruhe University of Applied Sciences. He completed his Bachelor's with the University of Cádiz in 2019, and his Master's with the University of Sevilla in 2021. His interests and research are focused on Federated Learning, Data Governance, Metadata Management, and Data Quality and Valuation.

TIM PIOTROWSKI is a PhD candidate at the Institute of Data-Centric Software Systems at Karlsruhe University of Applied Sciences, where his research focuses on the security of federated learning. He received his B.Sc. in Media and Communication Computer Science and his M.Sc. in Interactive Systems, with a focus on artificial intelligence, also from Karlsruhe University of Applied Sciences, Germany.

SASCHA GOHLKE is a PhD candidate at the Institute of Data Centric Software Systems at the Karlsruhe University of Applied Sciences. Prior to that, he completed a double-degree program with a Master's degree in Business Information Systems from the Karlsruhe University of Applied Sciences in Germany and a Master's degree in Computer Sciences from the Linnaeus University in Sweden.

ABDELHAFID ABOUAISSA received the BS degree from the Wroclaw University of Technology,Poland, in 1995, the MS degree from the University of Franche-Comte, Besançon, France, in 1996, the PhD degree from the University of Technology of Belfort-Montbéliard, France, in January 2000 and the accreditation to supervise research from the University of Haute-Alsace, France in 2017. He is a full professor in computer science at the University of Haute-Alsace, France, since 2019. His research interest include security, resource management in WSN, IoT, 5G networks, e-healt and Blockchain. He is author/co-author of 5 patents and 180 published international journals and conferences.

WILFRID AZAN is professor of management sciences in Lumière Lyon 2 University Within the Faculty of Economics and Management, he teaches MIS and qualitative methods. He obtained his PHD in the University of East Paris. He has published in Knowledge Management Research and Practice and technological forecasting and social change. He is invited professor at the J.W. Goethe of Frankfurt am Main. He is a member of Coactis and of the Beta Strasbourg, UMR CNRS 7522.

ISMAIL BENNIS In 2009, he holds a license degree in Mathematics and Computer Science at the Université Mohammed V, Rabat, Morocco. In 2011, he received a master's degree in computer networks and Telecommunications from the same university. In 2015, he holds his PhD degree under joint supervision between the Université Mohamed V, Rabat, Morroco, and the Université de Reims Champagne-Ardenne, France. Between 2015 and 2017 he works as a temporary professor for research and teaching (A.T.E.R) at University of Reims. After that, and until 2020, he works as an associate professor at La Rochelle Universiy. His research interests include the routing protocols with quality of service over the wireless sensor networks and IoT. Currently, he is an associate professor at University of Haute Alsace since September 2020.

ALFONSO GARCIA-DE-PRADO was born in Madrid, Spain, in 1972. He received the Ph.D. degree in Computer Science and Engineering from the University of Cadiz, Spain, in 2017. For several years, he has been a Programmer, an Analyst, and a Consultant for various international industry partners. Since 2011, he has been an Assistant Professor with the University of Cadiz. His research focuses on trending topics, such as CEP integration in service-oriented architectures, context awareness in the IoT, and its application in ambient assisted living.

MARC GILG is full professor of computer science at the University of Haute-Alsace since 2022. He defended his thesis in mathematics on May 19, 2000, then worked as a research engineer in computer science until 2006. At this date he becomes professor associate at IUT de Colmar, University of Haute-Alsace. He is member of IRIMAS laboratory. His research topics are wireless computer networks (ad hoc, sensor networks) and Cybersecurity. Marc GILG has co-authored 10 journals and 15 international conferences. Marc GILG is also head of Department Reseaux et Télécominications at IUT de Colmar.

EVA-MARIA NEUMANN received her B.Sc. and M.Sc. in Computer Science at the Karlsruhe Institute of Technology (KIT), Germany. Since 2022, she works as a software developer at INIT GmbH in Karlsruhe and supports the research department in selected projects.

ZOLTÁN NOCHTA received his Ph.D. degree in computer science from the University Karlsruhe, Germany, in 2004. He is chief development expert at SAP SE and has been working as part-time professor at Karlsruhe University of Applied Sciences since 2014. His research focuses on the security and practical applicability of federated machine learning in inter-company scenarios and on the integration of AI in business processes within various industry domains.

GUADALUPE ORTIZ Guadalupe Ortiz was born in Madrid, Spain in 1977. She obtained a Ph.D. degree in Computer Science from the University of Extremadura, Cáceres, Spain, in 2007. She is a Full Professor at the Computer Science and Engineering Department in the University of Cádiz, Spain. She has published over 100 peer-reviewed papers in international journals, workshops and conferences. Her research interests embrace context-awareness and stream processing in the scope of the IoT, with applications in ambient assisted living and smart cities, and a particular focus on the advancement and engineering of digital twins.

IOAN SZILAGYI is an AI and Data Scientist specialized in the development of AI-driven technologies and solutions across sectors such as cybersecurity, education, healthcare, and industry. He earned his PhD in Information and Communication Sciences from the University of Franche-Comté in 2014. His expertise spans the full spectrum of AI—from Symbolic AI, Ontologies and Semantic Web technologies to Machine Learning, Deep Learning, and Generative AI. He has also taught courses on Web Programming, Semantic Technologies, Internet of Things, and Artificial Intelligence. He currently serves as CTO of an AI-focused software company, where he leads the design and deployment of AI-based applications in multiple domains.

OLIVER P. WALDHORST (Member, IEEE) earned his diploma and Ph.D. in computer science from the University of Dortmund in 2000 and 2005 and his habilitation from Karlsruhe Institute of Technology in 2011. Since 2016, he has been Professor of Computer and Communication Networks at Karlsruhe University of Applied Sciences. His research focuses on communication protocols, distributed applications in mobile environments, and performance analysis of communication systems. He previously held research positions at Dortmund and Leipzig, was Senior Researcher at KIT (2006–2013) with stays at University of Toronto and TU Ilmenau, and worked on connected car communication at Daimler AG (2013–2016).

CHRISTIAN ZIRPINS Christian Zirpins received the Ph.D. degree in Computer Science from the University of Hamburg, Germany, in 2007. He is a Professor of Distributed Systems at the Faculty of Computer Science and Business Information Systems, Karlsruhe University of Applied Sciences (HKA), and the deputy spokesperson of the Institute of Data-Centric Software Systems (IDSS). He has authored more than 80 peer-reviewed papers in international journals, conferences, and workshops. His research interests include distributed middleware and software technologies for data-centric systems across IoT, edge, and cloud environments, with a focus on federated learning, edge intelligence, and resource-efficient orchestration of AI-enabled applications.