SSPRA: A Robust Approach to Continuous Authentication Amidst Real-world Adversarial Challenges

Frank (Sicong) Chen, Jingyu Xin, and Vir V. Phoha, *Fellow, IEEE*

Abstract—In real-world deployment, continuous authentication for mobile devices faces challenges such as intermittent data streams, variable data quality, and varying modality reliability. To address these challenges, we introduce a framework based on Markov process, named State-Space Perturbation-Resistant Approach (SSPRA). SSPRA integrates a two-level multi-modality fusion mechanism and dual state transition machines (STMs). This two-level fusion integrates probabilities from available modalities at each inspection (vertical-level) and evolves state probabilities over time (horizontal-level), thereby enhancing decision accuracy. It effectively manages modality disruptions and adjusts to variations in modality reliability. The dual STMs trigger appropriate responses upon detecting suspicious data, managing data fluctuations and extending operational duration, thus improving user experience. In our simulations, covering standard operations and adversarial scenarios like zero to non-zero-effort (ZE/NZE) attacks, modality disconnections, and data fluctuations, SSPRA consistently outperformed all baselines, including Sim's HMM and three state-of-the-art deep-learning models. Notably, in adversarial attack scenarios, SSPRA achieved substantial reductions in False Alarm Rate (FAR) - 36.31%, 36.58%, and 8.26% - and improvements in True Alarm Rate (TAR) - 33.15%, 33.75%, and 5.1% compared to the DeepSense, Siamese-structured network, and UMSNet models, respectively. Furthermore, it outperformed all baselines in modality disconnection and fluctuation scenarios underscores SSPRA's potential in addressing real-world challenges in mobile device authentication.

Index Terms—Continuous authentication, Multi-modality fusion, wearable devices, behavior biometrics, modality disconnection

1 Introduction

The widespread adoption of wearable and mobile technologies has significantly escalated the importance of robust continuous authentication systems. As individuals increasingly rely on these technologies for both personal and professional purposes, ensuring secure access to sensitive data through effective authentication becomes paramount. These devices generate substantial data, which, when effectively synthesized, can considerably bolster an individual's digital security.

Existing single-modality systems often suffer from inaccuracies and vulnerabilities [1], [2], [3]. Despite substantial research in multi-modal continuous authentication, the dynamic interplay between user verification and device-based systems remains complex challenges [4], [5], [6], [7], [8]. Predominant challenges include device reliability, the sporadic connectivity of modalities [9], [10] — a common challenge in real-world applications, and fluctuations in data quality, which critically impact the accuracy and reliability of authentication. Moreover, many existing authentication frameworks neglect the sequential continuity of data and temporal dependencies between authentication attempts [11], [12], which results in high false positives and struggles with inconsistent data quality.

In response to these challenges, we propose a framework named State-Space Perturbation-Resistant Approach (SSPRA). SSPRA is predicated on the first-order Markov process, where the system's current state is determined by its immediately preceding state combined with newly acquired information. SSPRA incorporates a two-level multi-modality fusion: the 'vertical level' utilizes all available modalities obtained at current state, while the 'horizontal level' considers the temporal influence of previous states. Figure 1 in Section 3.2 demonstrates multi-modality continuous authentication in a real-world context, illustrating these levels. At each system check, multiple simultaneous observations from various modalities (vertical level) are combined with prior state decisions (horizontal level) to ascertain the current state. Moreover, we introduce a 'Suspense' state and two state transition machines (STMs), operational under different conditions depending on whether suspicious data is detected (see Figure 3 in Section 3.2). This dual STM integration enhances system resilience, minimizes unnecessary disruptions, and extends the continuous authentication system's operational duration.

We simulated four real-world scenarios to evaluate SSPRA in continuous gait-based authentication: standard operation, adversarial attacks, modality disruption, and data fluctuation. We crafted a threat model for the adversarial attacks, detailed in Section 4.3. It encompasses various combinations of zero-effort (ZE) and non-zero-effort (NZE) attacks targeting different modalities. Our comparative analysis with four baseline models, including Sim *et al.*'s HMM-based verification system [13] and three state-of-the-art deep learning frameworks, DeepSense [14], Siamese-structure Network [15], and Transformer-based UMSNet [16], highlights SSPRA's superior performance. In adversarial attack conditions, SSPRA significantly outperforms

Manuscript received 12 Novemeber 2023; revised 14 January 2024; accepted 19 February 2024.
The authors are with the College of Engineering and Computer Science, Syracuse University, Syracuse, New York, 13244, U.S.A. (email: schen154@syr.edu, jxin05@syr.edu, vvphoha@syr.edu)

deep learning baselines with an 80.48% True Alarm Rate (TAR), surpassing DeepSense, Siamese-structured Network, and UMSNet by 33.15%, 33.75%, and 5.1%, respectively. Its False Alarm Rate (FAR) of only 16.79% is also notably lower than that of the baselines. In temporary modality disconnection scenarios, SSPRA's FAR is up to a quarter of that seen in Sim's HMM-based model. During tests with data fluctuation, SSPRA maintains the lowest FAR and the highest True Pass Rate (TP_aR). Particularly with 20% Gaussian noise, its FAR is only approximately one-ninth of DeepSense's. Additionally, in standard operation, SSPRA extends operational time by about 50% compared to baselines. These results demonstrate SSPRA's resilience and operational efficacy, highlighting its robustness in continuous authentication systems.

1.1 Main Contributions

With the introduction of the SSPRA, we advance the domain of continuous authentication by addressing critical limitations in current authentication models, particularly those encountered during real-world deployment. These include challenges such as maintaining authentication capabilities during modality disconnections, adapting to varying data quality, and preserving system continuity without sacrificing security. Our specific contributions are as follows:

1) **Consideration of System and Data Continuity:** Unlike previous continuous authentication models that perform isolated one-time inspections and overlook the continuity of data and systems [7], [17], SSPRA emphasizes this aspect by integrating past and current states, thereby enhancing the system's operational duration.
2) **Resilience to Modality Disconnection:** Our SSPRA uniquely addresses the issue of disconnected modalities — a scenario often ignored by existing work [11], [18] — by reliably operating with the available data, rather than generating synthetic data [19], [20], thus maintaining authentication integrity, especially when cross-modality correlation is not clearly defined.
3) **Improved Handling of Data Quality Variation:** Incorporating dual state transition machines (STMs), we introduce a 'Suspense' state into SSPRA to serve as a buffer against erratic data quality, increasing the model's resilience to noise, reducing unnecessary interruptions, and improving user experience.
4) **Adaptable Modality Reliability Handling:** In contrast to models that use fixed weighting schemes for modality reliability [5], [12], SSPRA dynamically adjusts weights based on real-time assessments of each modality's performance, ensuring a more robust and context-aware authentication process.
5) **Flexible Modality Management:** SSPRA's design supports post-deployment modifications, such as the inclusion or exclusion of modalities, offering a level of customization in user authentication that is not commonly provided by existing models [21], [22].

The rest of the paper is structured as follows: Section 2 delineates prior work and outlines the challenges in multi-modal fusion and authentication systems. Section 3 delves into the architecture of SSPRA, explicating the two-level fusion process, dual STMs, and the operational mechanisms. The evaluation metrics, experimental setup, and threat model are detailed in Section 4. Section 5 presents a thorough analysis of SSPRA's performance in a gait-based authentication scenario using mobile devices. Finally, Section 6 consolidates our findings and discusses their implications for the advancement of continuous authentication systems.

2 Motivation and Prior Work

In this section, we discuss the evolution of authentication systems, spotlighting the shift from single-modality to multi-modality fusion approaches and the inherent challenges in deploying these systems in real-world scenarios. We then introduce our SSPRA model, highlighting its novel contributions in addressing these challenges, and provide a comparison with a closely related study to underscore the advancements we bring to the field.

2.1 Advancements in Single-Modality and Multi-Modality Fusion Approaches

Existing studies have explored single-modality authentication systems [6], [23], [24], [25]. Despite their simplicity, these systems often suffer from limited accuracy [26], [27], [28] and are susceptible to adversarial attacks [29], [30], [31]. This has led to a growing interest in multi-modality systems, which aim to enhance security by integrating multiple modalities.

Multi-modality fusion approaches can be categorized into feature-level, decision-level, and score-level fusion. In feature-level fusion, features or raw data from multiple modalities are combined, as seen in the work of Pham *et al*. [22], who extracted features from three modalities for a breath-based authentication system. Yoon *et al*. introduced a pretrained implicit-ensemble transformer model for open-set authentication using touchstrokes and gait signals [32], and Delgado-Santos *et al*. [33] introduced M-GaitFormer for gait verification, which utilizes a Transformer to extract features from multiple gait signals. Decision-level fusion integrates decisions from separate classifiers by using rules like the maximum rule [5] and majority voting rule [12]. Score-level fusion involves normalizing and fusing scores from individual classification models to arrive at a final decision, with various techniques such as sum, product, maximum, and minimum rules [34], [35], [36]. Applications of these rules are seen in the work of Sitová *et al*. [37], who used the weighted sum rule for three modalities, and Ray-Dowling *et al*. [38], who evaluated score-level fusion on behavioral biometric-based continuous user authentication.

Each fusion level mentioned presents its own set of advantages and challenges. Feature-level fusion offers rich data insights but relies heavily on the availability of all modalities. Decision-level fusion provides operational independence for each modality, yet it can suffer from information loss about modality reliability. Score-level fusion offers a balance by allowing operation despite modality unavailability and offering detailed data integration. While it demands more complex normalization and higher computational resources, we selected it for SSPRA due to its

effectiveness in capturing comprehensive modality-specific details, which is crucial for enhancing multi-modal continuous authentication accuracy.

2.2 Advancements and Challenges in Multi-Modality Continuous Authentication Systems

The integration of multiple modalities in continuous authentication systems provides a comprehensive approach to user verification in dynamic real-world environments. Fridman *et al*. [5] combined keystroke dynamics with mouse movement data, evaluating each sensor's efficacy by analyzing false acceptance and rejection metrics for perpetual authentication. Buriro *et al*. fused actions before answering a call with voice recognition at the score-level for unobtrusive behavioral user authentication in [39]. In [40], they integrated these actions at the feature-level to propose a continuous user authentication mechanism. Deb *et al*. [41] proposed a Siamese Long Short-Term Memory (LSTM) network to extract deep temporal features from data generated from smartphones such as keystroke dynamics and GPS location, and introduced a passive user authentication method on smartphones utilizing these data. Lastly, Ray-Dowling *et al*. [42] provided a survey on stationary mobile behavioral biometrics, focusing on motion sensors and supporting non-motion, sporadic modalities like swipes and keystrokes.

While multi-modality systems show clear potential, deploying them in real-world scenarios poses various challenges. A major concern is the sporadic disconnection of devices due to unforeseen factors like Bluetooth interference or network issues. These disconnections impede the system's data access from the affected modality, making the fusion model non-functional. One approach that researchers have explored to tackle this is the imputation of missing data. Generally, these methods include matrix completion techniques and autoencoder-based strategies. While matrix completion methods [43], [44] often assume data is missing at random, this assumption may not hold true for continuous authentication where data omissions can span extended durations. Autoencoder-based methods leverage available modalities to impute missing data. Du *et al*. [45] introduced a semi-supervised multi-modal variational autoencoder to generate incomplete emotional data for emotional recognition, positing that emotional data from an individual exhibits inherent inter-modality correlations. Tsai *et al*. [19] developed a model that discerns and generates data from modality-specific representations. Ma *et al*. [20] presented SMIL, designed for scenarios with significant modality absence. Instead of relying on available modalities, their reconstruction network estimates missing data by computing a weighted sum of modalities. All these methods presuppose the existence of cross-modality correlations. However, unlike multi-modal benchmarks like CMU-MOSI [46] and AV-MNIST [47], where correlations are evident between modalities, the relationship between certain biometrics might be ambiguous or non-existent, potentially varying across datasets. This ambiguity complicates the evaluation of imputed data quality, with inaccuracies potentially compromising classification outcomes.

Another challenge lies in overlooking the continuous nature of data acquisition and analysis. Traditional continuous authentication approaches often treat each inspection as a discrete event [48], [49], neglecting the inherent temporal dependencies and historical context within continuous data streams. Such an approach is susceptible to transient data fluctuations. Additionally, while continuous monitoring with integrated anomaly detection has found applications in domains like electricity distribution [50] and autonomous vehicles [51], it has not been widely applied in continuous authentication. Furthermore, not all low-quality or fluctuating data series are necessarily anomalies. Nonetheless, if not addressed appropriately, these data can compromise the efficacy of the model.

2.3 How SSPRA Overcomes Identified Limitations

Given the limitations of single-modality systems and the challenges associated with multi-modality continuous authentication systems, coupled with the need to recognize the continuous nature of data streams, we introduce our solution: State-Space Perturbation-Resistant Approach (SSPRA). This model utilizes a two-level multi-modality fusion mechanism based on score-level fusion with dual STMs, emphasizing the system's continuity and navigating real-world challenges.

The vertical-level fusion effectively manages scenarios where certain modalities or sensors may sporadically disconnect due to connectivity problems or Bluetooth congestion. Unlike many existing models that either overlook this issue or impute missing data using other modalities, our approach strictly uses only available modalities. This guarantees that the system stays operational and provides precise decisions even when modalities are not highly correlated. We also directly evaluate how our model performs in these situations. Additionally, our design allows for easy addition or removal of modalities after system setup, enabling customization to meet individual requirements. Horizontal fusion emphasizes the system's continuity and the natural temporal patterns in signals, enhancing the overall accuracy and reliability of the fusion process.

Notably, the "Suspense" state in our STMs serves as a buffer between normal operation and alarm triggering, significantly reducing false alarms from short-lived low-quality or fluctuating data—a situation frequently neglected in previous studies but addressed in our experiments.

Comparison with Sim *et al.* 's work [13]

Sim *et al*. 's work [13] is closely related to our study, where they introduced a multi-modality continuous verification system based on the hidden Markov model (HMM). However, they assumed that only one observation could be obtained at each inspection. Conversely, our model operates under the **simultaneous data assumption**, positing that observations from diverse modalities, devices, or sensors are concurrently available, even if they are received within a brief interval. Moreover, our model goes beyond just having a "Normal" and an "Alert" state. We **integrate a "Suspense" state and utilizes dual STMs** for varied scenarios, enhancing its ability to manage inconsistent and fluctuating data. Lastly, our model provides **multiple user-configurable parameters**, allowing system designers to balance between minimizing false alarms and swiftly detecting abnormalities, tailored to specific application needs. These distinctions highlight the enhancements of our model over their work.

To further demonstrate the improved performance, we have also implemented Sim's HMM as a baseline model in Section 5.4 for comparative analysis with our SSPRA.

3 State-Space Perturbation-Resistant Approach (SSPRA) for Continuous Authentication

3.1 Backgrounds

3.1.1 Markov Assumptions and Rationale for their Application in Continuous Biometric Authentication Systems

A Markov chain or Markov process [52] is a stochastic model that describes a sequence of possible events (s_t) ordered by time ($t = 1, 2, ...$). The core principles of all Markov models are based on two key assumptions. First, the **limited horizon assumption** posits that the likelihood of occupying a specific state at time t is determined exclusively by the state at time $t-1$. The **stationary process assumption**, the second assumption, posits that the conditional probability distribution for the subsequent state, based on the present state, stays consistent over time.

In continuous biometric authentication systems, these Markovian assumptions are especially relevant. The **limited horizon assumption** aligns well with how authentication states evolve, typically influenced by the most recent state. For instance, in gait-based authentication, the likelihood of a specific gait pattern at any moment is closely related to the immediate past sequence of movements. Similarly, the **stationary process assumption** is applicable, as biometric data tend to exhibit stable patterns over time, with only minor fluctuations. These principles provide a robust framework for modeling the dynamics in continuous authentication systems, which is foundational to our SSPRA design.

3.1.2 SSPRA Notation

TABLE 1: Notation Used in the SSPRA

Symbol	Definition
P_1	State transition machine 1, activated when no suspicious data is detected
P_2	State transition machine 2, activated when suspicious data is detected
s_0 ="Normal"	Initial state of the model
$M_0 = \{\}$	Initial observations of the model
Ω_1 = {"Normal", "Suspense"}	State spaces for P_1
Ω_2 = {"Normal", "Suspense", "Alert"}	State spaces for P_2
$\mathcal{S} = \{s_0, s_1, s_2, \ldots\}$	Event space, where each event represents a system state at time t
$\mathcal{M}_t = \bigcup_{i=1}^{t} M_i$	Matching scores obtained up to time t
$M_t = \{m_{t1}, m_{t2}, \ldots, m_{tk_t}\}$	Matching scores of all available modalities obtained at time t
k_t	Number of available observations at time t

In Table 1, we present a list of notations utilized in the SSPRA formulation. These notations include symbols representing state transition machines (P_1 and P_2), potential system states ("Normal", "Suspense", and "Alert"), and sets denoting events and accumulated matching scores from observations (e.g., $\mathcal{S}$, $\mathcal{M}_t$). Each symbol in the table is integral to our methodology, and the following sections will further elucidate how these symbols are utilized in the SSPRA.

3.2 Determination of Each Modality's Likelihood: Matching Scores and Probability Mass Functions (PMFs)

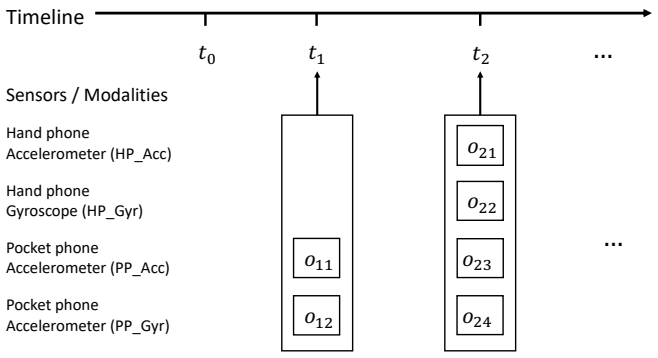


Fig. 1: Example illustrating simultaneous observations from multiple modalities/sensors influencing the system's state.

SSPRA incorporates a two-level fusion strategy: vertical-level, merging simultaneous observations from different modalities, and horizontal-level, integrating past and present states. Figure 1 showcases simultaneous data capture across modalities, where observations within a brief window are considered synchronized due to potential synchronization delays in real-world scenario.

This section focuses on calculating the likelihood for each modality based on system-acquired observations. In SSPRA, each modality operates independently, under the assumption that different modalities, utilizing varied sensors, capture distinct subject attributes and are largely uncorrelated or show ambiguous or hard-to-define correlations. This independence not only simplifies the fusion process by eliminating the need to account for covariance but also enhances system reliability, ensuring that issues in one modality do not impact others.

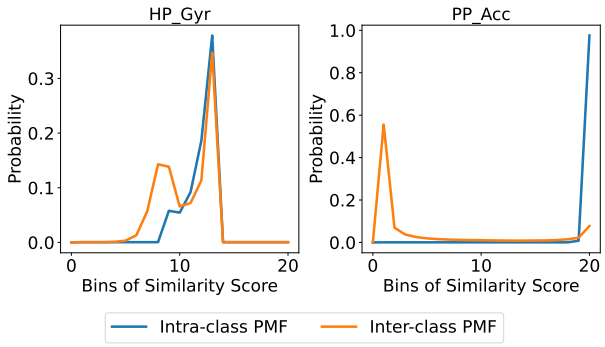


Fig. 2: PMFs of matching scores from two gait-based classifiers. The left plot represents a poorly performing classifier (HP_Gyr), while the right one depicts a well-performing classifier (PP_Acc). The data for these PMFs are sourced from the SU-AIS BB-MAS dataset [53].

Each modality's independent classifier generates a matching score (m_{tk_t}) that compares captured signals with

the genuine subject's template. Instead of directly merging these scores, SSPRA employs a probability mass function (PMF) approach, calculating PMFs for both intra- and inter-class samples based on training data. These PMFs represent the likelihood of obtaining specific matching scores for each class ($P(m_{tj}|\text{intra-class})$ and $P(m_{tj}|\text{inter-class})$), as displayed in Figure 2. For example, a matching score of 0.6 from the classifier shown in Figure 2 on the left corresponds to conditional probabilities of $P(\text{score} = 0.6|\text{intra-class}) = 0.38$ and $P(\text{score} = 0.6|\text{inter-class}) = 0.35$, indicating the probabilities of this score occurring within intra-class (genuine subject) and inter-class (other subjects) distributions, respectively.

Although matching scores are inherently continuous, the practicality of creating an accurate probability density function (PDF) is limited by the availability of training samples. Therefore, SSPRA treats scores as discrete values and employs the binning method to construct PMFs, providing a pragmatic approximation of the PDF.

3.3 Two-level Multi-modality Fusion Method: Combining Current Observation Likelihoods with Past State Likelihoods

In this section, the titles of each equation within the textboxes (in bold) serve as interactive links. Clicking on these titles redirects to the corresponding numerical examples in Appendix C, which demonstrate the practical application of each equation in real-world scenarios.

Given the matching scores ($M_t = \{m_{t1}, m_{t2}, ..., m_{tk_t}\}$) from k_t observations obtained at time t and the likelihood of each possible observation ($P(m_{tj}|s_t)$, where $j \in (1, 2, ..., k_t)$) derived from PMFs, the vertical-level fusion calculates the conditional probability $P(M_t|s_t)$ by employing the product rule to fuse the likelihoods of all available modalities obtained at time t:

Vertical-level fusion (fusing all available modalities)

$$P(M_t|s_t) = P(m_{t1}, m_{t2}, ..., m_{tk_t}|s_t) = \prod_{i=1}^{k_t} P(m_{ti}|s_t) \tag{1}$$

By applying the Law of Total Probability, we sum over all possible states at time $t-1$ to calculate the likelihood of the past state. Based on the **stationary process assumption**, the conditional probability of transition from previous state s_{t-1} to the current state t ($P(s_t|s_{t-1})$) is defined in the state transition machines. Thus, the likelihood of the past state is computed as follows:

Calculating likelihood of past state

$$P(s_t|\mathcal{M}_{t-1}) = \sum_{s_{t-1} \in \Omega_1 \text{or } \Omega_2} P(s_t|s_{t-1}) \cdot P(s_{t-1}|\mathcal{M}_{t-1}) \tag{2}$$

In line with the **limited horizon assumption**, the likelihood of the past state ($P(s_t|\mathcal{M}_{t-1})$) encapsulates all pertinent information up to time $t-1$. The current state is inferred based on this likelihood and the new observations at time t. Consequently, the horizontal-level fusion amalgamates the past state likelihoods with the vertical-level fusion outcomes using Bayes' rule, to deduce the probability $P(s_t = \omega_i|\mathcal{M}_t)$ for each possible state (ω_i) at time t:

Horizontal-level fusion (combines past and present states)

$$P(s_t|\mathcal{M}_t) \propto P(M_t|s_t) \cdot P(s_t|\mathcal{M}_{t-1}) \tag{3}$$

The probability of each possible state at time t can then be normalized using equation (Eq. 4):

$$P(s_t = \omega_i|\mathcal{M}_t) = \frac{P(s_t = \omega_i|\mathcal{M}_t)}{\sum_{\omega_j \in \Omega_1 or \Omega_2} P(s_t = \omega_j|\mathcal{M}_t)} \tag{4}$$

The current system state is determined by the state with the highest probability:

$$s_t = \underset{\omega_i \in \Omega_1 \text{or } \Omega_2}{\arg\max}\ P(s_t = \omega_i|\mathcal{M}_t) \tag{5}$$

Addressing Zero Probabilities and Absence of Observations: To manage instances where $P(m_{ti}|s_t)$ might equal zero, which can occur due to limited training data or insufficient binning in PMFs, we introduce a small constant value in each bin, ensuring non-zero probabilities. In situations with no new observations at time t, the matching score set $\mathcal{M}_t$ defaults to $\mathcal{M}_{t-1}$. Thus, $P(s_t|\mathcal{M}_t)$ is calculated via Eq. (2), as $P(s_t|\mathcal{M}_t) = P(s_t|\mathcal{M}_{t-1})$. This feature allows our system to continue operating even if all modalities are unavailable.

3.3.1 *Addressing Varying Reliability Across Modalities*

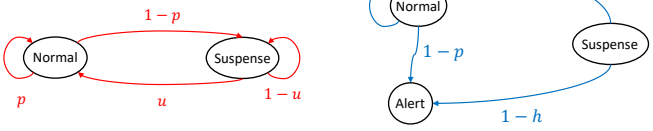
In SSPRA, the vertical-level fusion (Eq. (1)) adeptly manages the varying reliability and performance of different modalities through its PMF-based approach.

The PMFs serve as indicators of each modality's performance. Substantial overlap between the intra-class and inter-class PMFs typically signals weaker classifier performance, while minimal overlap indicates a more effective classifier, as depicted in Figure 2. Moreover, the impact of each modality's matching score on the fusion process is modulated based on its PMF. Underperforming modalities, characterized by comparable PMF values for different classes, exert less influence on Eq. (1), as their likelihood contributions are relatively uniform regardless of the class. In contrast, modalities with distinct PMF disparities between classes have a greater impact on the fusion outcome, ensuring that more reliable modalities play a more substantial role in the decision-making process.

For example, suppose at time t, the system obtains an observation from HP_Gyr (m_{t1}) and another observation from PP_Acc (m_{t2}), with similarity scores of 0.6 and 0.95, respectively. Utilizing PMFs as shown in Figure 2, we derive $P(m_{t1}|\text{intra-class}) = 0.38$, $P(m_{t1}|\text{inter-class}) = 0.35$, $P(m_{t2}|\text{intra-class}) = 0.98$, and $P(m_{t2}|\text{inter-class}) = 0.08$. At the vertical-level fusion (Eq. 1), we fuse these likelihoods: $P(M_t|\text{intra-class}) = 0.38 \times 0.98 = 0.3724$ and $P(M_t|\text{inter-class}) = 0.35 \times 0.08 = 0.028$. The calculation indicates that the poorly performing modality (HP_Gyr)

influences both $P(M_t|\text{intra-class})$ and $P(M_t|\text{inter-class})$ similarly due to its close likelihoods from intra-class and inter-class, respectively. In contrast, the better-performing modality (PP_Acc) affects these probabilities differently due to its divergent class likelihoods. This example illustrates how vertical-level fusion adeptly manages modalities with varying performance, demonstrating its effectiveness in the fusion process.

3.4 Dual State Transition Machines (STMs): Bolstering System Stability and Improving User Experience



(a) STM 1 (P_1) is activated in the absence of suspicious data.

(b) STM 2 (P_2) is activated upon the detection of suspicious data.

Fig. 3: Illustrations of dual STMs, where circles represent states and arrows denote transition probabilities between states.

SSPRA employs dual state transition machines (STMs) designed to navigate the complexities of low-quality and fluctuating data in real-world scenarios. Figure 3 illustrates these STMs, with circles representing states and arrows indicating transition probabilities between states. The absence of an arrow between two states signifies that there's no transition possible between them.

The "Normal" state signifies the system's standard operation without any detected attacks. The "Alert" state is activated upon detecting an anomaly, subsequently initiating an alarm. The "Suspense" state is indicative of potential abnormalities, leading the system to seek additional observations before deciding to transition to the "Alert" state. Thus, the first STM (P_1) operates in the absence of any suspicious data, while the second STM (P_2) comes into play upon detecting potential abnormalities. The transition probabilities in these STMs, as used in Eq. 2, represent $P(s_t|s_{t-1})$. For example, $P(s_t = \text{Normal}|s_{t-1} = \text{Normal}) = p$.

Two user-defined thresholds, $T_{\text{stay in normal}}$ and $T_{\text{back to normal}}$, play a crucial role in the transition process. $T_{\text{stay in normal}}$ determines the likelihood threshold for maintaining the 'Normal' state, while $T_{\text{back to normal}}$ is used to decide when to transition back to 'Normal' from the 'Suspense' state. How these criteria are applied within the SSPRA is detailed in Section 3.5.

3.4.1 Incorporation of Time-lapse Effect in STMs

SSPRA, recognizing the dynamic nature of real-world conditions, moves beyond the **stationary process assumption** of constant $P(s_t|s_{t-1})$. It employs exponential decay functions ($e^{-k\cdot\Delta_t}$) to represent the diminishing influence of past states over time, thus allowing the system to adapt to changing conditions and remain contextually relevant. In this model, k represents the decay factor and Δt is the time interval since the last inspection. For instance, to have the transition probability from 'Normal' to 'Normal' decrease to $1/3$ every 20 seconds, we set $k = -\ln(3)/20$, leading to a transition probability $p = e^{-\ln(3)/20\cdot\Delta t}$.

This methodology ensures SSPRA's decision-making is temporally sensitive, with decay factors typically defined by system designers based on domain expertise or analysis of historical data.

3.4.2 The Role of the "Suspense" State in Continuous Authentication

Traditional binary ('Normal' and 'Alert') continuous authentication systems often struggle with high false alarm rates due to transient data inconsistencies, such as low-quality observations or fluctuations, common in practical settings. Our SSPRA incorporates a 'Suspense' state into the STM design, enhancing this concept beyond merely providing an additional opportunity for data acquisition and decision-making. Upon detecting suspicious data, SSPRA triggers P_2, necessitating new observations from all available modalities and combining these with the previous decision. This design enhances system stability, minimizes false alarms, and improves the overall user experience by reducing unnecessary interruptions.

3.5 Operating Mechanism and Decision Making in SSPRA

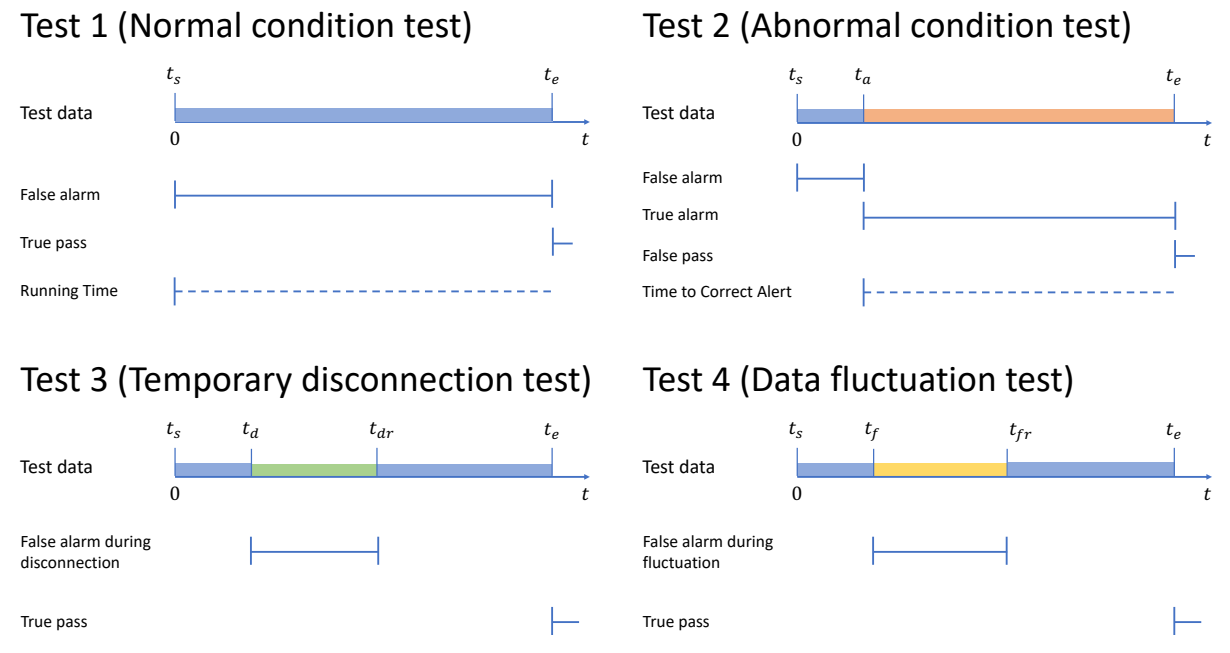
This section elucidates the operational and decision-making mechanisms of the SSPRA. Appendix C provides a more detailed walk-through with numerical examples.

During each system inspection, observations from all active modalities are gathered. The likelihood of these observations belonging to either intra-class or inter-class is then computed using the corresponding classification models and PMFs (as detailed in Section 3.2). This is followed by the determination of the likelihood for each potential system state, utilizing the two-level fusion methodology (Section 3.3) and the dual STMs (Section 3.4). Decision-making regarding state transitions is governed by user-defined thresholds: $T_{stay\ in\ normal}$ and $T_{back\ to\ normal}$.

In scenarios devoid of abnormalities, such as anomalous or poor-quality data, the system remains in the "Normal" state, given that $P(s_t = \text{Normal}) > T_{stay\ in\ normal}$. However, upon detecting suspicious data, the system transitions to the "Suspense" state at time t_1 if $P(s_{t_1} = \text{Normal}) < T_{stay\ in\ normal}$. Following this, the system obtains new observations at time t_2, and if $P(s_{t_2} = \text{Normal}) > T_{back\ to\ normal}$, the system reverts back to the "Normal" state. If persistent or genuine abnormalities are identified, the system escalates to the "Alert" state, thereby activating an alarm to signal potential attacks or spoofing attempts.

4 Experiment Configurations and Performance Evaluation Metrics

In this section, we first outline the performance metrics used to evaluate continuous authentication systems, and then we provide a detailed explanation of the four real-world scenarios we devised, including the specific configurations for each test, and the thread model designed for adversarial condition.



Test #	Test 1 (Normal condition test)	Test 2 (Abnormal condition test)	Test 3 (Temporary disconnection test)	Test 4 (Data fluctuation test)
Objective	Evaluate the system's performace under normal condition	Evaluate the system's efficiency and reponsiveness in detecting abnormal data	Evaluate the system's robustness and resilience in response to the temporal disconnection of certain sensors	Evaluate the system's robustness and resilience in response to the data fluctuation
Test Construction	The test data only consists of normal data from the monitored subject	After a fixed period of time (t_a), the rest of certain sensors are substituted by abnormal data (simulation of stress or attack)	After a fixed period of time(t_d), certain sensors are disconnected from the system for several fixed periods of time and reconnected to the system at t_{dr} (simulation of temporary disconnection)	After a fixed period of time (t_f), all sensors are added Gaussian noise for several fixed periods of time and removed at t_{fr} (simulation of data fluctuation)
Performace metrics	FAR, RT	FAR, TAR, FPR, TCA	FARDD, TPR	FARDF, TPR
Desired Outputs	The system should not generate any alarm, with low FAR and long RT	The system should generate an alarm as soon as the abnormal data occurs, with high TAR, low FAR, and short TCA	The system should not generate any alarm, with high TPR and low FARDD	The system should not generate any alarm, with high TPR and low FARDF

Fig. 4: Summary and graphical illustrations of the four tests used for evaluating the performance of SSPRA and baselines.

4.1 Performance metrics

To evaluate and compare SSPRA's performance with baseline models, we utilize both standard metrics like True Alarm Rate (TAR) and False Alarm Rate (FAR), as well as additional rate-based and time-based metrics. A summary of these metrics is presented in Table 2. Particularly, Reliable Running Time (RRT) is utilized to assess the system's usability, reflecting the duration a user can effectively use the system without interruptions, while Time to Correct Alarm (TCA) is used to measure the system's quickness in responding to abnormal data.

4.2 Experiment Configurations

In validating SSPRA's efficacy, we simulated four real-world scenarios within a continuous gait-based authentication system. Figure 4 details these tests, outlining objectives, structures, evaluation metrics, and expected outcomes. The graphical illustrations of each test above that table provide a clear visual representation of each test scenario.

Test 1 evaluates the system under normal operation using only genuine user data. In this scenario, the system is expected not to generate any alarms and to remain operational throughout the test, leading to a low FAR and extended RRT. **Test 2** focuses on SSPRA's response to various attack types, as detailed below in Section 4.3. **Test 3** simulates temporary modality disconnections by zeroing out data from certain modalities for set durations. The system is expected to detect these disconnections while continue operating during these periods until the end of the test, aiming for a low FARDD and a high TP_aR, given the absence of data from

TABLE 2: Summary of Metrics Used for Evaluating the Performance of Continuous Authentication Systems

Abbr.	Full Word	Definition / Calculation
Rate-based Metrics		
TP_aR	True Pass Rate	Number of alarm-free divided by total tests when no abnormal data is present.
FP_aR	False Pass Rate	Number of fail-to-alarm divided by total tests when abnormal data is present.
FARDD	False Alarm Rate During Disconnection	Number of false alarms occurring during any device disconnection divided by total tests.
FARDF	False Alarm Rate During Fluctuation	Number of false alarms occurring during periods of data fluctuation divided by total tests.
Time-based Metrics		
RRT	Reliable Running Time	Total duration from system initiation to termination in the absence of abnormal data.
TCA	Time to Correct Alarm	Time period between the onset of abnormal data and the system transitioning to the "Alert" state.

other subjects. Finally, **Test 4** introduces Gaussian noise into genuine users' data to emulate fluctuating data conditions. As the data are solely from genuine users, a system resilient to such fluctuations should not trigger false alarms and should maintain operation throughout the test, resulting in a low FARDF and high TP_aR.

TABLE 3: Specific Experimental Configurations for Each Test

Test	Configurations
All Tests	$t_s = 0s$, t_e = 45s
Test 1	t_e = 45s
Test 2	t_a = 10s
Test 3	t_d = 10s, $t_{dr} = t_d$ + 5s / 10s / 15s
Test 4	t_f = 10s, $t_{fr} = t_f$ + 5s / 10s / 15s

Table 3 details the experimental configurations for each test, including specifies values for t_e, t_a, t_d, t_{dr}, t_f, and t_{fr}, as delineated in Figure 4. For Tests 3 and 4, we selected three different durations to simulate modality disconnection and data fluctuation.

4.3 Threat Model for Abnormal Condition (Test 2)

To evaluate the resilience of SSPRA against adversarial attacks, we designed a threat model comprising two primary attack types: Zero-Effort and Non-Zero-Effort attacks.

1) **Zero-Effort (ZE) Attacks:** These attacks simulate scenarios in which an imposter tries to gain access by presenting their own data to the sensors, impersonating the genuine user.
2) **Non-Zero-Effort (NZE) Attacks:** In NZE attacks, the imposter tries to deceive the system by replaying the legitimate user's data to the sensor, representing a more sophisticated form of intrusion.

The threat scenarios are based on the assumption that attackers possess some familiarity with the authentication system, especially the sensors used, but lack knowledge about the specific models employed in the system.

We crafted four attack scenarios of varying difficulty, differentiated by the number of modalities subjected to ZE and NZE attacks. The simplest scenario entails no NZE attacks (equivalent to all modalities undergoing ZE attacks), wherein the genuine user's data is replaced with a randomly selected imposter's data to simulate the attack. As the difficulty increases, one, two, or three modalities are targeted with NZE attacks. This involves substituting the genuine user's data with their own data from a different sessions, while the remaining modalities are subjected to ZE attacks using an imposter's data. Utilizing the genuine user's own data in NZE attacks (like replay attacks) poses a significant challenge for the classifiers, particularly when multiple high-performing modalities are targeted, creating an arduous adversarial environment for the authentication system. A detailed discussion of SSPRA's performance in this test is illustrated in Section 5.4.

5 Substantiation of SSPRA through a Continuous Gait-based Authentication System

Continuous authentication in mobile devices, especially via gait analysis, strikes a balance between security and convenience. By leveraging built-in sensors such as accelerometers and gyroscopes, these systems authenticate users continuously through unique gait patterns. The non-replicable and effortless nature of gait-based authentication makes it an ideal application for our SSPRA.

5.1 Dataset Description and Data Pre-processing

For our experiment, the SU-AIS BB-MAS dataset [53], comprising gait data from 117 individuals, was utilized. This dataset, focusing on gait data captured by accelerometer and gyroscope from mobile device, is highly representative of real-world applications in gait-based authentication and verification [54], [55], [56].

The data collection involved two stages, where participants carried mobile devices equipped with three-axis accelerometers and gyroscopes, mimicking common usage patterns. In Stage 1, participants carried a tablet in their hand and a smartphone in their pocket, while in Stage 2, they held a smartphone in hand and another in their pocket. Each stage yielded approximately 90 seconds of gait data from four sensors operating simultaneously at a frequency of 100 Hz. The dual-device setup, involving both smartphone and tablet, provides a rich multi-modal dataset, ideal for evaluating the performance of our continuous gait-based authentication system across various sensor modalities and user scenarios.

During data preprocessing, we first filtered out subjects with incomplete or noisy data, resulting in a dataset of 96 subjects. Each subject's data was then split into two parts: one part, comprising 45 seconds of data, was designated for tests in the continuous authentication part. The other part was allocated for training, validation, and testing of each individual classification model and all deep-learning baseline models. The training set consisted of 8,000 samples, while the validation and testing sets each contained 1,000 samples, all balanced with an equal number of positive and negative samples.

5.2 Multi-modal Baseline Models

To benchmark our SSPRA, we implemented Sim's HMM (as discussed in Section 2), along with three state-of-the-art deep-learning multi-modal models in our experiments: the DeepSense-based model ("DeepSense"), Siamese-structured neural networks ("SiameseNet"), and Transformer-based UMSNet model ("UMSNet".)

DeepSense, developed by Yao *et al*. [14], employs ensembles of deep LSTM networks. It addresses challenges in Human Activity Recognition (HAR) using wearable devices, such as imbalanced and poor-quality data, common in real-life datasets. DeepSense's effectiveness in tasks like user identification through biometric motion analysis makes it a suitable baseline for comparing with SSPRA.

Siamese Networks, as implemented by Adel *et al*. [15], uses two identical deep neural networks containing Convolutional Neural Networks (CNN) and LSTM to learn the same embeddings from input signals and calculate distance between them. Its focus on individual authentication using inertial gait data aligns well with SSPRA's objectives in continuous gait-based authentication.

UMSNet, developed by Wang *et al*. [16], combines lightweight sensor residual blocks with the Transformer architecture to learn local and global multi-modal feature embeddings from multiple sensors. Its ability to be customized for various time series classification tasks makes it an apt model for benchmarking against our SSPRA.

Employing these models as baselines enables us to position SSPRA within the context of current state-of-the-art approaches, emphasizing efficiency, practicality, and adaptability — crucial aspects for real-world applications in continuous authentication systems.

5.3 Training and Testing Process

5.3.1 Training for SSPRA and Baseline Models

For both our SSPRA and Sim's HMM, the decision models do not require training but necessitate hyperparameter tuning, such as adjusting parameters in STMs and the transition thresholds. Nonetheless, pre-trained classifiers for each modality are essential. We adopted Siamese-structured neural networks for the gait-based classification in each modality. This choice is based on their established efficacy in inertial sensor-based authentication tasks [15], [31], [57]. Each Siamese network, trained for a specific sensor, was tasked with discriminating between genuine and non-genuine subjects' sensor data. The resulting similarity scores indicate the likelihood of data belonging to the genuine user. To ensure a fair comparison, the same classifiers were employed for both SSPRA and Sim's HMM. Classifiers are labeled for ease of reference: PP (phone in pocket), HT (tablet in hand), HP (phone in hand), with Acc (accelerometer) and Gyr (gyroscope) indicating the sensors. The F1-scores of each trained classifier are: PP_Acc (90.42%), PP_Gyr (85.43%), HT_Acc (74.34%), HT_Gyr (71.75%), HP_Acc (72.59%), and HP_Gyr (68.63%). Hyper-parameters used in our SSPRA and Sim's HMM are detailed in Appendix A.

Additionally, as detailed in Section 5.2, we employed three deep-learning models as baselines: DeepSense, SiameseNet, and UMSNet. Each model independently processes data from all four sensors to determine if an alarm should be generated. In terms of performance, DeepSense achieved F1-scores of 93.20% in Stage 1 and 91.65% in Stage 2. SiameseNet recorded scores of 85.55% in Stage 1 and 85.08% in Stage 2, while UMSNet attained F1-scores of 93.90% in Stage 1 and 92.56% in Stage 2.

5.3.2 Testing of the Continuous Gait-based Authentication System

For system testing, state inspections occur at intervals Δt, randomly set between 2 to 4 seconds. Each inspection involves capturing 2 seconds of data from all operational modalities. We conducted each test 20 times to average the performance metrics for SSPRA and the baseline models.

In Sim's HMM, given the limitation of accessing only one modality at a time, a random modality is chosen for each inspection. This approach prevents bias towards consistently selecting the highest-performing modality, which would otherwise reduce the system to a single-modality operation when all modalities are functional. It also ensures the validity of tests like modality disconnection scenarios, where the performance impact of disconnecting the least effective modality can be appropriately assessed.

5.4 Performance Evaluation in four Tests

In our experiment, SSPRA and the baseline models were evaluated using data from 96 subjects across both stages, Stage 1 (PP + HT) and Stage 2 (PP + HP), as detailed in Section 5.1. Our analysis revealed a marginal performance advantage in Stage 1 over Stage 2 for most models. This disparity is likely due to differences in sensor reliability between the tablet and the phone. As outlined in Section 5.3.1, classification models using data from the tablet held in hand (HT) generally exhibit better performance compared to those using data from the phone held in hand (HP).

Given our primary focus on comparing SSPRA's efficacy against the baseline models in varied test scenarios, we have opted to present and discuss only the results from Stage 1 in this section. This decision is aimed at providing a clear and focused analysis on the model performance, rather than the device-specific factors. Still, the detailed results for Stage2 are provided in tables in Appendix B.

TABLE 4: Test 1 (standard operation): Average FAR and RRT for SSPRA and baseline models.

Metrics	SSPRA	HMM	DeepSense	SiameseNet	UMSNet
FAR	**21.14%**	69.08%	68.32%	67.72%	54.24%
RRT (s)	**38.28**	25.02	24.33	23.82	28.13

Test 1 performance: 21.14% average FAR and 38.28 seconds RRT in 45-second tests - Test 1 focused on assessing SSPRA and baseline models under standard operation. Thus, each test can only result in either a true pass until the test duration ends (TP_aR) or a false alarm during the test duration (FAR), as described in Figure 4. Consequently, TP_aR and FAR are inversely related (TP_aR + FAR = 1) and we opted to focus on FAR as a measure of the system's propensity to generate false alarms under normal operation.

The results, illustrated in Table 4, highlight SSPRA's effectiveness. Despite the individual classifiers used in SSPRA showing less effectiveness than the all three deep-learning baseline models, SSPRA still surpassed these models in

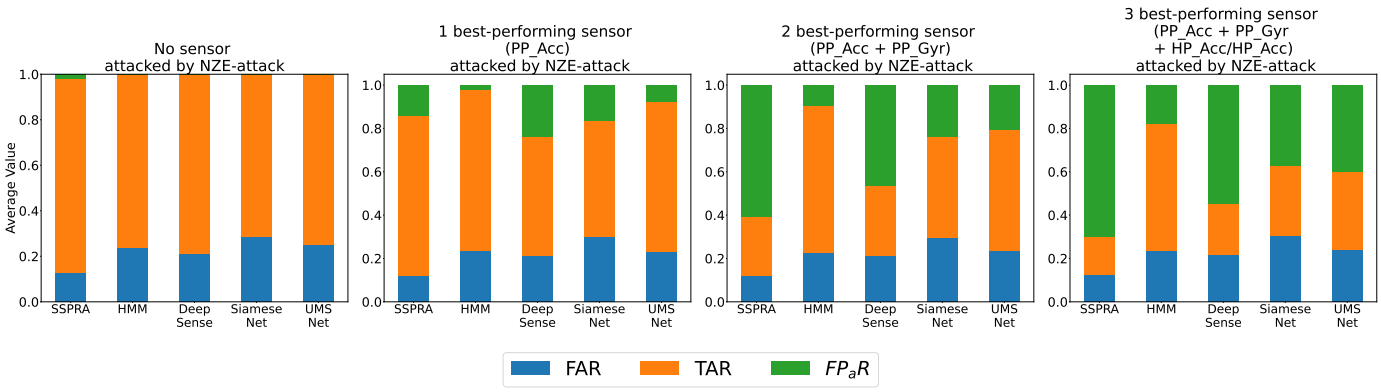


Fig. 5: Test 2 (under adversarial attacks): Average FAR, TAR, and FP_aR for SSPRA and baseline models under non-zero-effort (NZE) attacks on different numbers of modalities.

achieving a lower False Alarm Rate (FAR) and extended Reliable Running Time (RRT). For example, compared to UMSNet, which exhibited the best performance among the baselines, SSPRA achieved a 33.1% reduction in FAR and RRT by an additional 9.15 seconds.

The enhanced performance of SSPRA over deep-learning baseline models stems from its emphasis on system state continuity and the impact of previous states on the current one, aspects not considered in these baselines. Unlike deep-learning models that treat each system inspection as isolated, SSPRA's approach of incorporating the preceding state into current assessments effectively reduces false alarms and extends operational time. Furthermore, despite Sim's HMM's regard for system continuity, its limitation to a single modality per inspection constrains its analytical capabilities. This restriction makes it more prone to errors compared to SSPRA, which utilizes all available modalities for a more comprehensive analysis.

Test 2 performance: 80.27% TAR against ZE attacks and 4.24 seconds TCA - Test 2 was designed to assess system resilience against various threat scenarios, as detailed in Section 4.3. The performance metrics are detailed in Table 5, and Figure 5 visually contrasts SSPRA's performance with baseline models.

When all modalities face ZE attacks (equivalent to the absence of NZE attacks), SSPRA outperforms baseline models with an average TAR of 80.48%, exceeding Sim's HMM, DeepSense, and SiameseNet models by more than 30%, and surpassing UMSNet by 5.53%, as illustrated in Table 5a and Figure 5. This superior performance continues when one high-performing modality undergoes an NZE attack, with SSPRA improving TAR by up to 43.43% compared to the baselines. Notably, Sim's HMM, DeepSense, and SiameseNet models exhibit significantly high FAR in these scenarios, as shwon in Table 5b, suggesting a propensity for false alarms even prior to actual attacks. This also reflects their lesser efficiency in continuous authentication scenarios, as opposed to one-time authentication.

In our SSPRA, as discussed in Section 3.2, a fundamental assumption is that modalities operate independently. This assumption, however, presents challenges when multiple top-performing modalities are simultaneously subjected to Non-Zero Effort (NZE) attacks (attacked by genuine subject's data). In such scenarios, Sim's HMM achieved higher

TABLE 5: Test 2 (under adversarial attacks): Average TAR, FAR, and TCA for SSPRA and baseline models in Stage 1.

(a) Average TAR for SSPRA and baseline models.

NZE-attacked sensors	SSPRA	HMM	DeepSense	SiameseNet	UMSNet
No	**80.48%**	50.38%	46.90%	46.30%	74.95%
1 best-performing sensor (PP_Acc)	**71.47%**	46.52%	28.04%	31.68%	68.97%
1 worst-performing sensor (HT_Gyr)	**80.27%**	52.07%	46.63%	46.47%	75.22%
2 best-performing sensor (PP_Acc + PP_Gyr)	21.20%	39.40%	15.49%	27.55%	55.98%
2 worst-performing sensor (HT_Gyr + HT_Acc)	79.62%	49.62%	47.17%	45.49%	74.13%
3 best-performing sensor (PP_Acc + PP_Gyr + HT_Acc)	13.59%	30.43%	9.78%	13.21%	36.41%
3 worst-performing sensor (HT_Gyr + HT_Acc + PP_Gyr)	74.40%	47.45%	44.95%	44.57%	71.36%

(b) Average FAR for SSPRA and baseline models.

NZE-attacked sensors	SSPRA	HMM	DeepSense	SiameseNet	UMSNet
No	**16.79%**	48.91%	53.10%	53.37%	25.05%
1 best-performing sensor (PP_Acc)	**17.28%**	50.00%	52.66%	53.53%	23.21%
1 worst-performing sensor (HT_Gyr)	**16.09%**	47.17%	53.37%	53.10%	23.75%
2 best-performing sensor (PP_Acc + PP_Gyr)	17.61%	47.83%	53.42%	52.99%	23.48%
2 worst-performing sensor (HT_Gyr + HT_Acc)	16.41%	49.13%	52.72%	53.26%	23.10%
3 best-performing sensor (PP_Acc + PP_Gyr + HT_Acc)	16.68%	49.89%	52.72%	53.70%	23.70%
3 worst-performing sensor (HT_Gyr + HT_Acc + PP_Gyr)	16.25%	49.78%	53.26%	53.80%	22.61%

(c) Average TCA (s) for SSPRA and baseline models.

NZE-attacked sensors	SSPRA	HMM	DeepSense	SiameseNet	UMSNet
No	**4.23**	2.36	0.96	1.25	1.79
1 best-performing sensor (PP_Acc)	4.72	2.97	1.71	1.86	3.09
1 worst-performing sensor (HT_Gyr)	**4.24**	2.49	0.93	1.32	1.87
2 best-performing sensor (PP_Acc + PP_Gyr)	2.62	3.45	1.57	1.87	5.00
2 worst-performing sensor (HT_Gyr + HT_Acc)	4.24	2.41	1.02	1.40	2.40
3 best-performing sensor (PP_Acc + PP_Gyr + HT_Acc)	3.06	2.80	1.52	1.86	5.22
3 worst-performing sensor (HT_Gyr + HT_Acc + PP_Gyr)	4.38	3.50	1.08	1.43	3.38

TAR, likely due to our strategy of randomly selecting modalities, which might inadvertently favor modalities under Zero Effort (ZE) attacks that are easier to detect. UMSNet, on the other hand, demonstrated better performance than SSPRA, possibly due to its architecture and transformer module. Its network architecture not only allows it to learn from individual sensors but also to understand the relationships between them, thus aiding in anomaly detection when sensor data do not match a single user's profile.

Regarding TCA, SSPRA averaged 4.23 seconds under no

TABLE 6: Test 3 (modality disconnection): Average FARDD and TP_aR for SSPRA and baseline models with disconnection of one or two top or bottom performing modalities over short (5s), medium (10s), and long (15s) durations in Stage 1.

Disconnected sensors		1 best-performing sensor (PP_Acc)			1 worst-performing sensor (HT_Gyr)			2 best-performing sensor (PP_Acc + PP_Gyr)			2 worst-performing sensor (HT_Gyr + HT_Acc)		
Length of disconnection		Short	Medium	Long	Short	Medium	Long	Short	Medium	Long	Short	Medium	Long
FARDD	SSPRA	8.64%	12.23%	15.82%	2.23%	3.75%	4.67%	4.29%	8.75%	**12.66%**	2.50%	4.67%	5.82%
	HMM	15.49%	27.88%	41.41%	9.40%	16.58%	23.97%	17.07%	37.93%	57.17%	6.79%	11.63%	16.58%
TP_aR	SSPRA	72.39%	70.82%	68.10%	78.10%	78.64%	79.18%	73.75%	71.03%	**69.62%**	78.15%	77.66%	77.45%
	HMM	25.76%	22.61%	20.71%	32.39%	31.74%	33.10%	22.12%	13.80%	9.29%	35.49%	35.87%	39.84%

NZE attack and 4.24 seconds under an NZE attack scenarios, as detailed in Table 5c. The inclusion of the "Suspense" state in SSPRA presents a strategic design choice. While it effectively prevents premature alarms by requiring additional data verification, this process inherently extends the TCA. SSPRA's design necessitates at least two observation windows (4 seconds in our setup) for decision-making, compared to baseline models which may trigger alarms within a single observation window (2 seconds). This deliberate approach in SSPRA, favoring accuracy (TAR) and minimizing false alarms (FAR) over rapid detection, is evident in its substantial FAR reduction of over 36.58%. This trade-off prioritizes accurate and reliable detection, reducing unwarranted disruptions even if it means a slightly longer detection time, especially considering the potential risk to accuracy and increased FAR with faster decision times.

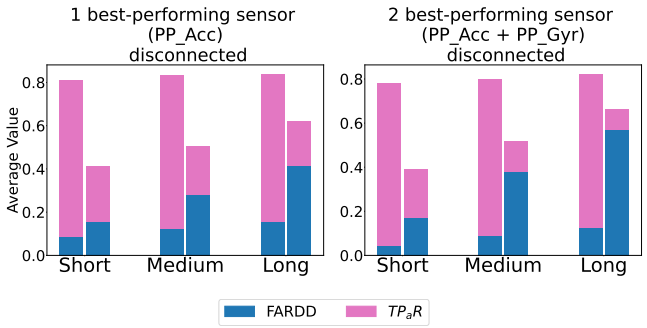


Fig. 6: Test 3 (modality disconnection): Average FARDD and TP_aR for SSPRA (left bars) and Sim's HMM (right bars) under disconnection of one or two best-performing modalities for short (5s), medium (10s), and long (15s) durations.

Test 3 Performance: substantially lower FAR (12.66%) during a 15-second disconnection of two best-performing modalities, compared to the baseline model (57.17%) - Test 3 was designed to assess SSPRA's resilience in the face of short (5s), medium (10s), and long (15s) modality disconnections. Due to the operational limitations of DeepSense, SiameseNet, and UMSNet, which require data from all four modalities simultaneously, this test's comparison is focused solely on Sim's HMM.

Figure 6 visualizes the numerical results from Table 6, comparing SSPRA with Sim's HMM. SSPRA consistently outperforms Sim's HMM, achieving lower False Alarm Rate During Disconnection (FARDD) and higher True Pass Rate (TP_aR) across all scenarios. This performance gap widens as disconnection duration increases and more top-performing modalities are disconnected. Notably, in the most challenging scenario with two top modalities disconnected for 15 seconds, SSPRA maintains a 69.62% TP_aR and only a 12.66% FARDD. In contrast, Sim's HMM's performance diminishes significantly, recording a mere 9.29% TP_aR and a high FARDD of 57.17%. Unlike Sim's HMM, which relies on a single, less reliable modality and suffers increased FARDD, SSPRA's use of all available modalities mitigates the impact of disconnection, maintaining a low FARDD and high TP_aR. These findings highlight SSPRA's robustness and reliability in managing temporary modality disconnections, a crucial factor for practical application.

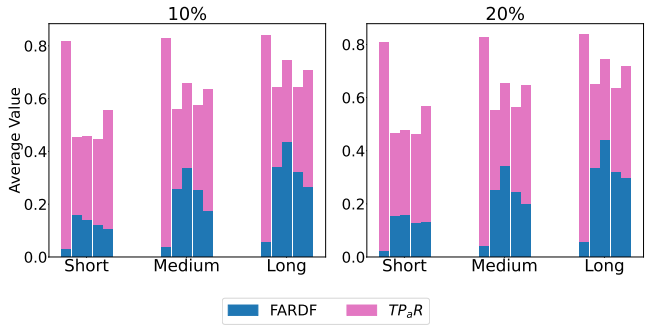


Fig. 7: Test 4 (data fluctuation): Average FARDF and TP_aR for SSPRA and baseline models under 10% or 20% Gaussian noise for short (5s), medium (10s), and long (15s) durations. Bars in each group, from left to right, represent SSPRA, Sim's HMM, DeepSense, SiameseNet, and UMSNet.

TABLE 7: Test 4 (data fluctuation): Average FARDF and TP_aR for SSPRA and baseline models across different levels and durations of data fluctuation in Stage 1.

(a) Average FARDF for SSPRA and baseline models.

Amount of Gaussian noise	Length of data fluctuation	SSPRA	HMM	DeepSense	SiameseNet	UMSNet
10%	Short	3.15%	15.76%	14.18%	12.34%	10.60%
	Medium	3.70%	26.09%	33.86%	25.54%	17.39%
	Long	5.87%	34.08%	43.80%	32.39%	26.68%
20%	Short	2.28%	15.54%	15.71%	12.77%	13.37%
	Medium	4.08%	25.43%	34.46%	24.51%	20.27%
	Long	**5.71%**	33.42%	44.18%	32.12%	29.73%

(b) Average TP_aR for SSPRA and baseline models.

Amount of Gaussian noise	Length of data fluctuation	SSPRA	HMM	DeepSense	SiameseNet	UMSNet
10%	Short	78.64%	29.51%	31.47%	32.50%	44.78%
	Medium	79.08%	30.00%	32.17%	32.12%	46.14%
	Long	78.32%	30.27%	30.87%	32.07%	44.29%
20%	Short	78.64%	30.92%	31.96%	33.26%	43.59%
	Medium	78.59%	29.78%	31.20%	31.85%	44.46%
	Long	77.93%	31.47%	30.22%	31.58%	41.96%

Test 4 Performance: Significant reduction in FARDF compared to baseline models - Test 4, designed to evaluate SSPRA's response to data fluctuations, is illustrated in Figure 7 with numerical results in Table 7. In all test scenarios, SSPRA consistently outperformed baseline models,

exhibiting the lowest False Alarm Rate During Fluctuation (FARDF) and the highest TP_aR. Notably, under extreme conditions with 20% Gaussian noise for 15 seconds, SSPRA maintained a FARDF of only 5.71%, which is one-fifth to one-eighth compared to all baselines. This performance is significantly better than all other baselines, with the largest difference observed against DeepSense, which registered the highest FARDF at 44.18%. This comparison underscores SSPRA's robustness and clearly demonstrates its enhanced resilience to data fluctuations, showcasing its potential in real-world applications where data quality can vary unpredictably.

SSPRA's robust performance against data fluctuations stems from its emphasis on system and signal continuity, combined with the strategic use of a 'Suspense' state. By recognizing that data are generally continuous with minimal abrupt changes over short periods, SSPRA effectively utilizes state continuity in its decision-making. This approach enhances resilience to sudden data variations. The 'Suspense' state further strengthens this capability, allowing additional time to gather observations and ensuring measured, accurate responses to suspicious data.

6 Conclusion and Discussion

In this paper, we present State-Space Perturbation-Resistant Approach (SSPRA), a novel solution devised to address the challenges prevalent in continuous authentication using mobile devices within dynamic real-world settings.Traditional models in this domain often fall short in managing intermittent data unavailability, data volatility, and the continuity of system states. SSPRA, by contrast, adeptly navigates these issues, offering a comprehensive and robust solution.

With the assumption that the authentication process adheres to first-order Markov process, our SSPRA implements a two-level fusion mechanism that combines the likelihood from multiple modalities (vertical-level fusion) and integrates the present and past states with time-lapsing effects (horizontal-level fusion). This two-level fusion bolsters the system's resilience during modality disconnections and facilitates the dynamic integration or exclusion of modalities. Moreover, SSPRA incorporates a "Suspense" state within its dual state transition machines to manage transient data quality issues, thereby minimizing disruptions and extending system operational time. This innovative approach also allows system designers to tailor specific suspension conditions for the system, particularly when critical modalities are disconnected.

Empirical evaluations in continuous gait-based authentication demonstrate SSPRA's superior performance. It consistently exhibits lower Fasle Alarm Rates (FARs) under standard operation and higher True Alarm Rates (TARs) against adversarial attacks compared to Sim's HMM and three state-of-the-art deep-learning models. Remarkably, SSPRA maintains its robustness in scenarios of modality disconnections and data fluctuations, as evidenced by its reduced FARs and improved True Pass Rates (TP_aRs). These results underscore SSPRA's adaptability and effectiveness in real-world scenarios.

While SSPRA has shown promising results, we recognize the inherent trade-off between minimizing FAR and achieving a swift Time To Correct Alart (TCA). Future research should focus on balancing these aspects more effectively. Efforts to refine individual classifier performance and explore alternative decay functions could further optimize SSPRA's accuracy and operational efficiency. Such advancements hold great promise for broadening SSPRA's application scope and enhancing its impact in various real-world continuous authentication contexts.

References

[1] P. K. Sahoo, H. K. Thakkar, W.-Y. Lin, P.-C. Chang, and M.-Y. Lee, "On the design of an efficient cardiac health monitoring system through combined analysis of ecg and scg signals," *Sensors*, vol. 18, no. 2, p. 379, 2018.

[2] P. Zontone, A. Affanni, R. Bernardini, A. Piras, and R. Rinaldo, "Stress detection through electrodermal activity (eda) and electrocardiogram (ecg) analysis in car drivers," in *2019 27th European Signal Processing Conference (EUSIPCO)*, pp. 1–5, IEEE, 2019.

[3] M. Shabaan, K. Arshid, M. Yaqub, F. Jinchao, M. S. Zia, G. R. Bojja, M. Iftikhar, U. Ghani, L. S. Ambati, and R. Munir, "Survey: smartphone-based assessment of cardiovascular diseases using ecg and ppg analysis," *BMC medical informatics and decision making*, vol. 20, pp. 1–16, 2020.

[4] A. Roy, T. Halevi, and N. Memon, "An hmm-based behavior modeling approach for continuous mobile authentication," in *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp. 3789–3793, IEEE, 2014.

[5] L. Fridman, A. Stolerman, S. Acharya, P. Brennan, P. Juola, R. Greenstadt, and M. Kam, "Multi-modal decision fusion for continuous authentication," *Computers & Electrical Engineering*, vol. 41, pp. 142–156, 2015.

[6] I. C. Stylios, O. Thanou, I. Androulidakis, and E. Zaitseva, "A review of continuous authentication using behavioral biometrics," in *Proceedings of the SouthEast European Design Automation, Computer Engineering, Computer Networks and Social Media Conference*, pp. 72–79, 2016.

[7] A. Mosenia, S. Sur-Kolay, A. Raghunathan, and N. K. Jha, "Caba: Continuous authentication based on bioaura," *IEEE Transactions on Computers*, vol. 66, no. 5, pp. 759–772, 2016.

[8] A. Krašovec, D. Pellarini, D. Geneiatakis, G. Baldini, and V. Pejović, "Not quite yourself today: Behaviour-based continuous authentication in iot environments," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 4, pp. 1–29, 2020.

[9] O. Dehzangi, M. Taherisadr, and R. ChangalVala, "Imu-based gait recognition using convolutional neural networks and multi-sensor fusion," *Sensors*, vol. 17, no. 12, p. 2735, 2017.

[10] J. Li and Q. Wang, "Multi-modal bioelectrical signal fusion analysis based on different acquisition devices and scene settings: Overview, challenges, and novel orientation," *Information Fusion*, vol. 79, pp. 229–247, 2022.

[11] S. Rasnayaka and T. Sim, "Action invariant imu-gait for continuous authentication," in *2022 IEEE International Joint Conference on Biometrics (IJCB)*, pp. 1–10, IEEE, 2022.

[12] Z. Shen, S. Li, X. Zhao, and J. Zou, "Mmauth: a continuous authentication framework on smartphones using multiple modalities," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 1450–1465, 2022.

[13] T. Sim, S. Zhang, R. Janakiraman, and S. Kumar, "Continuous verification using multimodal biometrics," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 4, pp. 687–700, 2007.

[14] S. Yao, S. Hu, Y. Zhao, A. Zhang, and T. Abdelzaher, "Deepsense: A unified deep learning framework for time-series mobile sensing data processing," in *Proceedings of the 26th international conference on world wide web*, pp. 351–360, 2017.

[15] O. Adel, M. Soliman, and W. Gomaa, "Inertial gait-based person authentication using siamese networks," in *2021 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–7, IEEE, 2021.

[16] J. Wang, H. Wei, Y. Wang, S. Yang, and C. Li, "Umsnet: An universal multi-sensor network for human activity recognition," *arXiv preprint arXiv:2205.11756*, 2022.

[17] R. Kumar, P. P. Kundu, D. Shukla, and V. V. Phoha, "Continuous user authentication via unlabeled phone movement patterns," in *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pp. 177–184, IEEE, 2017.

[18] Y. Cao, F. Li, H. Chen, X. liu, L. Zhang, and Y. Wang, "Guard your heart silently: Continuous electrocardiogram waveform monitoring with wrist-worn motion sensor," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 3, pp. 1–29, 2022.

[19] Y.-H. H. Tsai, P. P. Liang, A. Zadeh, L.-P. Morency, and R. Salakhutdinov, "Learning factorized multimodal representations," *arXiv preprint arXiv:1806.06176*, 2018.

[20] M. Ma, J. Ren, L. Zhao, S. Tulyakov, C. Wu, and X. Peng, "Smil: Multimodal learning with severely missing modality," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, pp. 2302–2310, 2021.

[21] H. Huang, P. Zhou, Y. Li, and F. Sun, "A lightweight attention-based cnn model for efficient gait recognition with wearable imu sensors," *Sensors*, vol. 21, no. 8, p. 2866, 2021.

[22] C. Pham, M.-H. Bui, V.-A. Tran, A. D. Vu, and C. Tran, "Personalized breath-based biometric authentication with wearable multimodality," *IEEE Sensors Journal*, vol. 23, no. 1, pp. 536–543, 2022.

[23] W. Louis, M. Komeili, and D. Hatzinakos, "Continuous authentication using one-dimensional multi-resolution local binary patterns (1dmrlbp) in ecg biometrics," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 12, pp. 2818–2832, 2016.

[24] Y. Yang, B. Guo, Z. Wang, M. Li, Z. Yu, and X. Zhou, "Behavesense: Continuous authentication for security-sensitive mobile apps using behavioral biometrics," *Ad Hoc Networks*, vol. 84, pp. 9–18, 2019.

[25] G. Dahia, L. Jesus, and M. Pamplona Segundo, "Continuous authentication using biometrics: An advanced review," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 10, no. 4, p. e1365, 2020.

[26] H. Locklear, S. Govindarajan, Z. Sitová, A. Goodkind, D. G. Brizan, A. Rosenberg, V. V. Phoha, P. Gasti, and K. S. Balagani, "Continuous authentication with cognition-centric text production and revision features," in *Ieee international joint conference on biometrics*, pp. 1–8, IEEE, 2014.

[27] S. Eberz, K. B. Rasmussen, V. Lenders, and I. Martinovic, "Evaluating behavioral biometrics for continuous authentication: Challenges and metrics," in *Proceedings of the 2017 ACM on Asia conference on computer and communications security*, pp. 386–399, 2017.

[28] K. Quintal, B. Kantarci, M. Erol-Kantarci, A. Malton, and A. Walenstein, "Contextual, behavioral, and biometric signatures for continuous authentication," *IEEE Internet Computing*, vol. 23, no. 5, pp. 18–28, 2019.

[29] C. Sousedik and C. Busch, "Presentation attack detection methods for fingerprint recognition systems: a survey," *Iet Biometrics*, vol. 3, no. 4, pp. 219–233, 2014.

[30] A. Buriro, B. Crispo, F. Delfrari, and K. Wrona, "Hold and sign: A novel behavioral biometrics for smartphone user authentication," in *2016 IEEE security and privacy workshops (SPW)*, pp. 276–285, IEEE, 2016.

[31] J. Xin, V. V. Phoha, and A. Salekin, "Combating false data injection attacks on human-centric sensing applications," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 2, pp. 1–22, 2022.

[32] J. Yoon, J. Park, K. Wagata, H. Park, and A. B. J. Teoh, "Pre-trained implicit-ensemble transformer for open-set authentication on multimodal mobile biometrics," in *Proceedings of the 31st ACM International Conference on Multimedia*, pp. 5909–5922, 2023.

[33] P. Delgado-Santos, R. Tolosana, R. Guest, R. Vera-Rodriguez, and J. Fierrez, "M-gaitformer: Mobile biometric gait verification using transformers," *Engineering Applications of Artificial Intelligence*, vol. 125, p. 106682, 2023.

[34] J. Kittler, M. Hatef, R. P. Duin, and J. Matas, "On combining classifiers," *IEEE transactions on pattern analysis and machine intelligence*, vol. 20, no. 3, pp. 226–239, 1998.

[35] M. He, S.-J. Horng, P. Fan, R.-S. Run, R.-J. Chen, J.-L. Lai, M. K. Khan, and K. O. Sentosa, "Performance evaluation of score level fusion in multimodal biometric systems," *Pattern Recognition*, vol. 43, no. 5, pp. 1789–1800, 2010.

[36] M. Hariri and S. B. Shokouhi, "Robustness of multi biometric authentication systems against spoofing," *Computer and Information Science*, vol. 5, no. 1, p. 77, 2012.

[37] Z. Sitová, J. Šeděnka, Q. Yang, G. Peng, G. Zhou, P. Gasti, and K. S. Balagani, "Hmog: New behavioral biometric features for continuous authentication of smartphone users," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 5, pp. 877–892, 2015.

[38] A. Ray-Dowling, D. Hou, S. Schuckers, and A. Barbir, "Evaluating multi-modal mobile behavioral biometrics using public datasets," *Computers & Security*, vol. 121, p. 102868, 2022.

[39] A. Buriro, B. Crispo, F. Del Frari, J. Klardie, and K. Wrona, "Itsme: Multi-modal and unobtrusive behavioural user authentication for smartphones," in *Technology and Practice of Passwords: 9th International Conference, PASSWORDS 2015, Cambridge, UK, December 7–9, 2015, Proceedings 9*, pp. 45–61, Springer, 2016.

[40] A. Buriro, B. Crispo, and M. Conti, "Answerauth: A bimodal behavioral biometric-based user authentication scheme for smartphones," *Journal of information security and applications*, vol. 44, pp. 89–103, 2019.

[41] 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 *2019 international conference on biometrics (ICB)*, pp. 1–8, IEEE, 2019.

[42] A. Ray-Dowling, D. Hou, and S. Schuckers, "Stationary mobile behavioral biometrics: A survey," *Computers & Security*, vol. 128, p. 103184, 2023.

[43] M. Qin, Z. Du, F. Zhang, and R. Liu, "A matrix completion-based multiview learning method for imputing missing values in buoy monitoring data," *Information Sciences*, vol. 487, pp. 18–30, 2019.

[44] J. Bai and S. Ng, "Matrix completion, counterfactuals, and factor analysis of missing data," *Journal of the American Statistical Association*, vol. 116, no. 536, pp. 1746–1763, 2021.

[45] C. Du, C. Du, H. Wang, J. Li, W.-L. Zheng, B.-L. Lu, and H. He, "Semi-supervised deep generative modelling of incomplete multi-modality emotional data," in *Proceedings of the 26th ACM international conference on Multimedia*, pp. 108–116, 2018.

[46] A. Zadeh, R. Zellers, E. Pincus, and L.-P. Morency, "Mosi: multimodal corpus of sentiment intensity and subjectivity analysis in online opinion videos," *arXiv preprint arXiv:1606.06259*, 2016.

[47] V. Vielzeuf, A. Lechervy, S. Pateux, and F. Jurie, "Centralnet: a multilayer approach for multimodal fusion," in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pp. 0–0, 2018.

[48] H. Feng, K. Fawaz, and K. G. Shin, "Continuous authentication for voice assistants," in *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, pp. 343–355, 2017.

[49] L. Azhari and A. M. Barmawi, "Activity attribute-based user behavior model for continuous user authentication,"

[50] G. Fenza, M. Gallo, and V. Loia, "Drift-aware methodology for anomaly detection in smart grid," *IEEE Access*, vol. 7, pp. 9645–9657, 2019.

[51] F. Van Wyk, Y. Wang, A. Khojandi, and N. Masoud, "Real-time sensor anomaly detection and identification in automated vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 1264–1276, 2019.

[52] J. R. Norris, *Markov chains*. No. 2, Cambridge university press, 1998.

[53] A. K. Belman, L. Wang, S. S. Iyengar, P. Sniatala, R. Wright, R. Dora, J. Baldwin, Z. Jin, and V. V. Phoha, "Su-ais bb-mas (syracuse university and assured information security - behavioral biometrics multi-device and multi-activity data from same users) dataset," 2019.

[54] D. Gafurov, E. Snekkenes, and P. Bours, "Gait authentication and identification using wearable accelerometer sensor," in *2007 IEEE workshop on automatic identification advanced technologies*, pp. 220–225, IEEE, 2007.

[55] T. T. Ngo, Y. Makihara, H. Nagahara, Y. Mukaigawa, and Y. Yagi, "The largest inertial sensor-based gait database and performance evaluation of gait-based personal authentication," *Pattern Recognition*, vol. 47, no. 1, pp. 228–237, 2014.

[56] A. H. Johnston and G. M. Weiss, "Smartwatch-based biometric gait recognition," in *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pp. 1–6, IEEE, 2015.

[57] C. Wu, X. Li, F. Zuo, L. Luo, X. Du, J. Di, and Q. Zeng, "Use it-no need to shake it! accurate implicit authentication for everyday objects with smart sensing," *Proceedings of the ACM on Interactive,*

Mobile, Wearable and Ubiquitous Technologies, vol. 6, no. 3, pp. 1–25, 2022.

APPENDIX A
HYPER-PARAMETERS IN SSPRA AND SIM'S HMM IN EXPERIMENTS

As delineated in Section 3.4, our SSPRA model encompasses six hyper-parameters in two STMs: p and u for P_1, p and h for P_2, along wit two state transitioning thresholds $T_{stay\ in\ normal}$ and $T_{back\ to\ normal}$. In contrast, Sim's HMM only uses a singular hyper-parameter, p, for its STM. For the experiments, six hyper-parameter configurations were tuned for SSPRA and four for Sim's HMM, and only the results from the optimal hyper-parameter sets are discussed in Section 5.4. The optimal sets for each model are as follows:

1) SSPRA: P_1: $p = e^{-\ln 2/10\cdot\Delta t}$, $u = e^{-\ln 2/4\cdot\Delta t}$; P_2: $p = e^{-\ln 2/10\cdot\Delta t}$, $h = e^{-\ln 2/4\cdot\Delta t}$; $T_{stay\ in\ normal} = 0.5$; $T_{back\ to\ normal} = 0.6$.
2) HMM: $p = e^{-\ln 2/8\cdot\Delta t}$.

APPENDIX B
TABULATED RESULTS FROM CONTINUOUS GAIT-BASED AUTHENTICATION SYSTEM

B.1 Tabulated Results for Test 1 from Stage 2

Refer to Section 5.4 and the discussion in **Test 1** for insights related to the results presented in Table 8.

TABLE 8: Test 1 (standard operation): Average FAR and RT for SSPRA and baseline models.

Metrics	SSPRA	HMM	DeepSense	SiameseNet	UMSNet
FAR	27.34%	74.73%	67.50%	66.96%	59.02%
RT (s)	36.90	23.02	24.39	22.67	26.20

B.2 Tabulated Results for Test 2 from Stage 2

Refer to Section 5.4 and the discussion in **Test 2** for insights related to the results presented in Table 9, 10, and 11.

TABLE 9: Test 2 (under adversarial attacks): Average TAR for SSPRA and baseline models in Stage 2.

NZE-attacked sensors	SSPRA	HMM	DeepSense	SiameseNet	UMSNet
No	78.32%	47.55%	49.24%	44.78%	72.55%
1 best-performing sensor (PP_Acc)	69.89%	45.43%	29.29%	31.30%	65.71%
1 worst-performing sensor (HT_Gyr)	77.55%	46.79%	49.29%	45.22%	71.36%
2 best-performing sensor (PP_Acc + PP_Gyr)	28.37%	40.05%	18.64%	28.59%	50.71%
2 worst-performing sensor (HT_Gyr + HT_Acc)	76.96%	46.85%	49.46%	43.86%	69.78%
3 best-performing sensor (PP_Acc + PP_Gyr + HT_Acc)	19.78%	32.93%	11.20%	10.98%	33.59%
3 worst-performing sensor (HT_Gyr + HT_Acc + PP_Gyr)	73.86%	46.03%	47.07%	45.16%	66.58%

B.3 Tabulated Results for Test 3 from Stage 2

Refer to Section 5.4 and the discussion in **Test 3** for insights related to the results presented in Table 12 and 13.

B.4 Tabulated Results for Test 4 from Stage 2

Refer to Section 5.4 and the discussion in **Test 4** for insights related to the results presented in Table 14 and 15.

TABLE 10: Test 2 (under adversarial attacks): Average FAR for SSPRA and baseline models in Stage 2.

NZE-attacked sensors	SSPRA	HMM	DeepSense	SiameseNet	UMSNet
No	19.24%	51.90%	50.71%	55.05%	27.45%
1 best-performing sensor (PP_Acc)	19.46%	52.07%	52.50%	55.27%	26.36%
1 worst-performing sensor (HT_Gyr)	19.13%	52.61%	50.71%	54.57%	28.64%
2 best-performing sensor (PP_Acc + PP_Gyr)	18.75%	51.30%	51.20%	54.29%	28.15%
2 worst-performing sensor (HT_Gyr + HT_Acc)	19.29%	52.39%	50.27%	54.84%	27.34%
3 best-performing sensor (PP_Acc + PP_Gyr + HT_Acc)	19.46%	51.52%	52.12%	54.62%	28.48%
3 worst-performing sensor (HT_Gyr + HT_Acc + PP_Gyr)	18.59%	50.49%	51.09%	53.70%	26.96%

TABLE 11: Test 2 (under adversarial attacks): Average TCA for SSPRA and baseline models in Stage 2.

NZE-attacked sensors	SSPRA	HMM	DeepSense	SiameseNet	UMSNet
No	4.11	2.06	1.09	1.48	1.69
1 best-performing sensor (PP_Acc)	4.39	2.82	1.57	2.32	2.99
1 worst-performing sensor (HT_Gyr)	4.01	2.08	1.01	1.46	1.67
2 best-performing sensor (PP_Acc + PP_Gyr)	3.28	3.08	1.50	2.40	4.10
2 worst-performing sensor (HT_Gyr + HT_Acc)	4.13	2.39	1.10	1.51	2.34
3 best-performing sensor (PP_Acc + PP_Gyr + HT_Acc)	3.01	3.02	1.74	2.00	4.58
3 worst-performing sensor (HT_Gyr + HT_Acc + PP_Gyr)	4.43	3.20	1.22	1.69	3.17

APPENDIX C
NUMERICAL ILLUSTRATION OF SSPRA OPERATIONAL PROCESS USING THE SU-AIS BB-MAS DATASET [53]

In this section, we illustrate the operation of SSPRA using a simplified version of the continuous gait-based verification system from Section 5. This example employs two sensors, handphone's gyroscope (HP_Gyr) and pocketphone's accelerometer (PP_Acc), as opposed to the four sensors used in the full system. Figure 2 provides the PMF plots of matching scores for these sensors. To streamline the calculation process, state abbreviations are as follows: "N" for "Normal", "S" for "Suspense", and "A" for "Alert". The parameters utilized are consistent with those reported in Appendix A.

C.1 At time $t_0 = 0\ s$: Legitimate User Log-in

Assume the legitimate user logs in at t_0 using their password and starts using the phone. The continuous authentication system initiates operation, and its state is set to "N". Notably, the system does not require observations from any sensors at this juncture.

Therefore, we obtain: $M_{t_0} = \{\}$, $\mathcal{M}_{t_0} = \{\}$, $P(s_{t_0} = \text{N}|\mathcal{M}_{t_0}) = 1$, $P(s_{t_0} = \text{S}|\mathcal{M}_{t_0}) = 0$.

C.2 At time $t_1 = 2.1\ s$: First Inspection

Assume the system acquires observations from HP_Gyr and PP_Acc at $t_1 = 2.1\ s$, with $\Delta t_1 = t_1 - t_0 = 2.1\ s$.

For each observation, the system processes it through the classifier, yielding matching scores $M_{t_1} = \{m_{t_1 1} = 0.6, m_{t_1 2} = 0.97\}$. Upon checking these with their PMFs, we ascertain $P(m_{t_1 1} = 0.6|s_{t_1} = \text{N}) = 0.38$, $P(m_{t_1 1} = 0.6|s_{t_1} = \text{S}) = 0.35$, $P(m_{t_1 2} = 0.97|s_{t_1} = \text{N}) = 0.98$, $P(m_{t_1 2} = 0.97|s_{t_1} = \text{S}) = 0.07$.

Applying Eq. (1), we fuse the likelihoods of each available modality at the vertical level:

TABLE 12: Test 3 (modality disconnection): Average FARDD and TP_aR for SSPRA and baseline models with disconnection of one top or bottom performing modalities over different durations in Stage 2.

Disconnected sensors		1 best-performing sensor (PP_Acc)			1 worst-performing sensor (HP_Gyr)		
Length of disconnection		Short (5s)	Medium (10s)	Long (15s)	Short (5s)	Medium (10s)	Long (15s)
FARDD	MMCMM	7.50%	13.37%	17.61%	1.25%	1.96%	3.21%
	HMM	13.59%	28.80%	43.15%	9.84%	16.47%	27.07%
TP_aR	MMCMM	69.13%	67.39%	63.21%	74.02%	73.75%	75.87%
	HMM	22.88%	19.29%	17.28%	26.03%	27.55%	28.59%

TABLE 13: Test 3 (modality disconnection): Average FARDD and TP_aR for SSPRA and baseline models with disconnection of two top or bottom performing modalities over different durations in Stage 2.

Disconnected sensors		2 best-performing sensor (PP_Acc + PP_Gyr)			2 worst-performing sensor (HP_Gyr + HP_Acc)		
Length of disconnection		Short (5s)	Medium (10s)	Long (15s)	Short (5s)	Medium (10s)	Long (15s)
FARDD	MMCMM	5.98%	10.38%	14.73%	1.52%	4.02%	4.51%
	HMM	17.45%	39.02%	53.70%	6.41%	12.72%	20.49%
TP_aR	MMCMM	68.59%	67.77%	65.22%	72.77%	74.02%	75.71%
	HMM	20.22%	13.26%	8.70%	30.76%	30.87%	34.67%

$$P(M_{t_1}|s_{t_1} = \mathrm{N}) = 0.38 \times 0.98 = 0.372,$$
$$P(M_{t_1}|s_{t_1} = \mathrm{S}) = 0.35 \times 0.07 = 2.45 \times 10^{-2}$$

Note: This calculation is derived directly from applying Eq. (1) from Section 3.3 to the given scenario.

Considering the system is in the "N" state, we use state transition machine P_1 to compute the likelihood of the previous state. Applying equation (2), we derive:

$$\begin{aligned} P(s_{t_1} = \mathrm{N}|\mathcal{M}_{t_0}) &= \sum_{s_{t_0} \in \Omega_1} P(s_{t_1} = \mathrm{N}|s_{t_0}) \cdot P(s_{t_0}|\mathcal{M}_{t_0}) \\ &= P(s_{t_1} = \mathrm{N}|s_{t_0} = \mathrm{N}) \cdot P(s_{t_0} = \mathrm{N}|\mathcal{M}_{t_0}) \\ &\quad + P(s_{t_1} = \mathrm{N}|s_{t_0} = \mathrm{S}) \cdot P(s_{t_0} = \mathrm{S}|\mathcal{M}_{t_0}) \\ &= e^{-\ln(2)/10\times 2.1} \times 1 + e^{-\ln(2)/4\times 2.1} \times 0 \\ &= 0.871 \end{aligned}$$

$$\begin{aligned} P(s_{t_1} = \mathrm{S}|\mathcal{M}_{t_0}) &= \sum_{s_{t_0} \in \Omega_1} P(s_{t_1} = \mathrm{S}|s_{t_0}) \cdot P(s_{t_0}|\mathcal{M}_{t_0}) \\ &= P(s_{t_1} = \mathrm{S}|s_{t_0} = \mathrm{N}) \cdot P(s_{t_0} = \mathrm{N}|\mathcal{M}_{t_0}) \\ &\quad + P(s_{t_1} = \mathrm{S}|s_{t_0} = \mathrm{S}) \cdot P(s_{t_0} = \mathrm{S}|\mathcal{M}_{t_0}) \\ &= (1 - e^{-\ln(2)/10\times 2.1}) \times 1 + (1 - e^{-\ln(2)/4\times 2.1}) \times 0 \\ &= 0.129 \end{aligned}$$

Note: The above calculations are derived directly from applying Eq. (2) from Section 3.3 to the given scenario.

Finally, employing equation (3), we fuse the likelihood of the previous state and current observations at the horizontal level:

$$\begin{aligned} P(s_{t_1} = \mathrm{N}|\mathcal{M}_{t_1}) &\propto P(M_{t_1}|s_{t_1} = \mathrm{N}) \cdot P(s_{t_1} = \mathrm{N}|\mathcal{M}_{t_0}) \\ &= 0.372 \times 0.871 = 0.324 \end{aligned}$$

$$\begin{aligned} P(s_{t_1} = \mathrm{S}|\mathcal{M}_{t_1}) &\propto P(M_{t_1}|s_{t_1} = \mathrm{S}) \cdot P(s_{t_1} = \mathrm{S}|\mathcal{M}_{t_0}) \\ &= 2.45 \times 10^{-2} \times 0.129 = 3.16 \times 10^{-3} \end{aligned}$$

TABLE 14: Test 4 (data fluctuation): Average FARDF for SSPRA and baseline models across different levels and durations of data fluctuation in Stage 2.

Amount of Gaussian noise	Length of data fluctuation	SSPRA	HMM	DeepSense	SiameseNet	UMSNet
10%	Short (5s)	2.77%	15.38%	11.03%	6.90%	8.70%
	Medium (10s)	6.68%	26.68%	27.93%	15.54%	21.90%
	Long (15s)	9.29%	36.63%	37.01%	21.63%	25.87%
20%	Short (5s)	2.77%	15.11%	11.68%	6.74%	11.36%
	Medium (10s)	6.96%	24.29%	28.10%	16.68%	24.89%
	Long (15s)	8.64%	35.82%	37.28%	22.12%	30.05%

TABLE 15: Test 4 (data fluctuation): Average TP_aR for SSPRA and baseline models across different levels and durations of data fluctuation in Stage 2.

Amount of Gaussian noise	Length of data fluctuation	SSPRA	HMM	DeepSense	SiameseNet	UMSNet
10%	Short (5s)	74.08%	28.48%	32.34%	32.12%	41.25%
	Medium (10s)	73.15%	26.25%	33.80%	34.18%	39.73%
	Long (15s)	73.10%	27.55%	33.10%	32.01%	40.22%
20%	Short (5s)	73.42%	27.66%	32.28%	32.50%	39.67%
	Medium (10s)	72.61%	28.97%	34.08%	33.86%	37.61%
	Long (15s)	73.26%	27.61%	32.12%	32.72%	36.68%

Note: The above calculations are derived directly from applying Eq. (3) from Section 3.3 to the given scenario.

After normalizing the two values above, we attain the likelihood of each state using equation (4): $P(s_{t_1} = \mathrm{N}) = 0.324/(0.324 + 3.16 \times 10^{-3}) = 0.991$

Since $P(s_{t_1} = \mathrm{N}) > T_{stay\ in\ normal}$, the system's state at t_1 is "N", indicating no suspicious data present.

C.3 At time $t_2 = 4.3\ s$: Second Inspection

At time $t_2 = 4.3\ s$, the system acquires observations from both sensors, yielding $\Delta t_2 = t_2 - t_1 = 2.2\ s$.

The derived matching scores from these observations are: $M_{t_2} = \{m_{t_2 1} = 0.5, m_{t_2 2} = 0.06\}$. Checking with their corresponding PMFs, we find $P(m_{t_2 1} = 0.5|s_{t_2} = \mathrm{N}) = 0.1$, $P(m_{t_2 1} = 0.5|s_{t_2} = \mathrm{S}) = 0.08$, $P(m_{t_2 2} = 0.06|s_{t_2} = \mathrm{N}) = 0.005$, and $P(m_{t_2 2} = 0.06|s_{t_2} = \mathrm{S}) = 0.08$.

We first execute the vertical fusion of these observations using Eq. (1): $P(M_{t_2}|s_{t_2} = \mathrm{N}) = 0.1 \times 0.005 = 5 \times 10^{-4}$, $P(M_{t_2}|s_{t_2} = \mathrm{S}) = 0.08 \times 0.08 = 0.064$

Then, we proceed with the horizontal fusion of the present observations and prior states using Eqs. (2) and (3): $P(s_{t_2} = \mathrm{N}|\mathcal{M}_{t_2}) \propto 5 \times 10^{-4} \times (e^{-\ln(2)/10\times 2.2} \times 0.991 + e^{-\ln(2)/4\times 2.2} \times 0.009) = 4.3 \times 10^{-4}$, $P(s_{t_2} = \mathrm{S}|\mathcal{M}_{t_2}) \propto 0.064 \times ((1 - e^{-\ln(2)/10\times 2.2}) \times 0.991 + (1 - e^{-\ln(2)/4\times 2.2}) \times 0.009) = 8.7 \times 10^{-4}$.

After normalizing the above likelihoods, we obtain $P(s_{t_2} = \mathrm{N}) = 0.331$ and $P(s_{t_2} = \mathrm{S}) = 0.669$.

As $P(s_{t_2} = \mathrm{N}) < T_{stay\ in\ normal}$, the system flags a suspicious data point at this stage. This could be due to low-quality or fluctuating data from any sensors, potential abnormalities, or possible sensor attacks. To confirm whether an alarm should be triggered, the system transitions to the "S" state at this point, awaiting a new set of observations from all available sensors.

C.4 At time $t_3 = 6.3\ s$: Third Inspection

Since the system transitioned to the "S" state during the last inspection, it waits for 2 seconds (equivalent to the duration

of the observation window) to collect observations from all accessible sensors. These observations will then aid the system in deciding whether to transition into the "A" state and trigger an alarm or revert back to the "N" state. Here, $\Delta t_3 = t_3 - t_2 = 2\ s$.

To illustrate the operational process of our MMCMM in both the aforementioned scenarios, we postulate that the system receives different observations in each scenario. It should be noted that, given the system transitioned to the "S" state at time t_2, we must use the state transition machine P_2 to compute the likelihood of each state in both scenarios.

C.4.1 Scenario: System Receives Matching Scores $M_{t_3} = \{m_{t_3 1} = 0.4, m_{t_3 2} = 0.22\}$

In this scenario, the probabilities are computed as: $P(m_{t_3 1} = 0.4|s_{t_3} = \text{N}) = 0.06$, $P(m_{t_3 1} = 0.4|s_{t_3} = \text{A}) = 0.14$, $P(m_{t_3 2} = 0.22|s_{t_3} = \text{N}) = 0.004$, $P(m_{t_3 2} = 0.22|s_{t_3} = \text{A}) = 0.018$.

The fusion at the vertical level is calculated using Eq. (1): $P(M_{t_3}|s_{t_3} = \text{N}) = 0.06 \times 0.004 = 2.4 \times 10^{-4}$, $P(M_{t_3}|s_{t_3} = \text{A}) = 0.14 \times 0.018 = 2.52 \times 10^{-3}$.

The fusion at the horizontal level is computed using Eqs. (2) and (3): $P(s_{t_3} = \text{N}|\mathcal{M}_{t_3}) \propto 2.4 \times 10^{-4} \times (e^{-\ln(2)/10\times 2} \times 0.331 + e^{-\ln(2)/2\times 2} \times 0.667) = 1.49 \times 10^{-4}$, $P(s_{t_3} = \text{A}|\mathcal{M}_{t_3}) \propto 2.52 \times 10^{-3} \times ((1 - e^{-\ln(2)/10\times 2}) \times 0.331 + (1 - e^{-\ln(2)/2\times 2}) \times 0.667) = 1.51 \times 10^{-3}$.

After normalizing the probabilities, we obtain: $P(s_{t_3} = \text{N}) = 0.135$, and $P(s_{t_3} = \text{A}) = 0.865$. Given that $P(s_{t_3} = \text{N}) < T_{back\ to\ normal}$, the system will transition to the "A" state, and an alarm will be triggered.

In this scenario, the data collected at time t_2 is highly likely indicative of an attack. The system identifies a potential threat at time t_2, verifies it at time t_3, and ultimately triggers an alarm following verification at time t_3.

C.4.2 Scenario: System Receives Matching Scores $M_{t_3} = \{m_{t_3 1} = 0.62, m_{t_3 2} = 0.98\}$

In this scenario, the probabilities are computed as: $P(m_{t_3 1} = 0.62|s_{t_3} = \text{N}) = 0.38$, $P(m_{t_3 1} = 0.62|s_{t_3} = \text{A}) = 0.35$, $P(m_{t_3 2} = 0.98|s_{t_3} = \text{N}) = 0.98$, $P(m_{t_3 2} = 0.98|s_{t_3} = \text{A}) = 0.07$.

The fusion at the vertical level is calculated using Eq. (1): $P(M_{t_3}|s_{t_3} = \text{N}) = 0.38 \times 0.98 = 0.372$, $P(M_{t_3}|s_{t_3} = \text{A}) = 0.35 \times 0.07 = 2.45 \times 10^{-2}$

The fusion at the horizontal level is computed using Eqs. (2) and (3): $P(s_{t_3} = \text{N}|\mathcal{M}_{t_3}) \propto 0.372 \times (e^{-\ln(2)/10\times 2} \times 0.331 + e^{-\ln(2)/2\times 2} \times 0.667) = 0.231$, $P(s_{t_3} = \text{A}|\mathcal{M}_{t_3}) \propto 2.45\times 10^{-2} \times ((1-e^{-\ln(2)/10\times 2}) \times 0.331 + (1-e^{-\ln(2)/2\times 2}) \times 0.667) = 9.41 \times 10^{-3}$.

After normalizing the probabilities, we obtain: $P(s_{t_3} = \text{N}) = 0.961$, and $P(s_{t_3} = \text{A}) = 0.039$. Since $P(s_{t_3} = \text{N}) > T_{back\ to\ normal}$, the system will return to the "N" state. In the next inspection, the system will use the state transition machine P_1 to compute the likelihood of each state, and it can utilize $P(s_{t_3} = \text{N}) = 0.961$ and $P(s_{t_3} = \text{S}) = 0.039$ as the likelihoods of each past state at time t_3.

In this scenario, the data collected at time t_2 is very likely to be low-quality or fluctuating data. The system transitions to the "S" state at time t_2 but reverts to the "N" state at time t_3 after verification that it is not under attack. In this manner, the system avoids the false alarm that might occur in conventional continuous monitoring systems.

C.5 At Time $t_4 = 8.5\ s$: Fourth Inspection

Suppose at time t_3, the system transitions back to the normal state, similar to the second scenario discussed earlier. Now, at time $t_4 = 8.5\ s$, the connection to HP_Gyr might get lost due to reasons such as connection issues. As such, the system only receives an observation from PP_Acc, which has a matching score of $M_{t4} = \{m_{t_4 1} = 0.97\}$. We can get $P(m_{t_4 1} = 0.97|s_{t_3} = \text{N}) = 0.98$ and $P(m_{t_4 1} = 0.97|s_{t_3} = \text{S}) = 0.07$ from its PMF.

As only one observation is available, we can skip the vertical-level fusion and directly calculate horizontal-level fusion using Eqs. (2) and (3) as follows: $P(s_{t_4} = \text{N}|\mathcal{M}_{t_4}) \propto 0.98 \times (e^{-\ln(2)/10\times 2.2} \times 0.961 + e^{-\ln(2)/2\times 2.2} \times 0.039) = 0.835$, $P(s_{t_4} = \text{S}|\mathcal{M}_{t_4}) \propto 0.07 \times ((1 - e^{-\ln(2)/10\times 2.2}) \times 0.961 + (1 - e^{-\ln(2)/2\times 2.2}) \times 0.039) = 0.010$.

After normalizing these probabilities, we can get $P(s_{t_4} = \text{N}) = 0.988$. Since $P(s_{t_4} = \text{N}) > T_{stay\ in\ normal}$, the system will remain in the "N" state. This scenario demonstrates the functioning of our MMCMM when certain modalities disconnect from the system or when the system fails to receive observations from certain modalities.

Frank (Sicong) Chen received his M.S. degree in Computer Science from Syracuse University in 2020 and is currently pursuing his Ph.D. in the Computer and Information Science and Engineering program at Syracuse University. His research interests encompass deep learning, security, biometrics, and human-computer interactions.

Jingyu Xin earned M.Sc. degrees in Electrical Engineering and Engineering Data Analytics & Statistics from Washington University in St. Louis in 2018. Presently, he is a Ph.D. candidate in the Computer and Information Science and Engineering program at Syracuse University, focusing on research in the security and robustness of ubiquitous computing applications.

Vir V. Phoha (M'96–SM'02–Fellow) received the Ph.D. degree in computer science from Texas Tech University, Lubbock, in 1992. He is currently a Professor of Electrical Engineering and Computer Science in the College of Engineering and Computer Science at Syracuse University. His research interest includes attack-averse authentication, optimized attack formulation, machine learning, anomaly detection, spatial-temporal pattern detection and event recognition, and knowledge discovery and analysis. Professor Phoha is a Fellow of AAAS; IEEE; NAI; SDPS; and is an ACM Distinguished Scientist.