Behavioral Biometrics for Continuous Authentication in the Internet of Things Era: An Artificial Intelligence Perspective

Yunji Liang, Sagar Samtani, Bin Guo, and Zhiwen Yu

Abstract—In the Internet of things (IoT) era, user authentication is an essential to ensure the security of connected devices and the customization of passive services. However, conventional knowledge-based and physiological biometric-based authentication systems (e.g., password, face recognition, and fingerprints) are susceptible to shoulder surfing attacks, smudge attacks and heat attacks. The powerful sensing capabilities of IoT devices including smartphones, wearables, robots and autonomous vehicles enable the continuous authentication (CA) based on behavioral biometrics. Artificial Intelligence (AI) approaches hold significant promise in sifting through large volumes of heterogeneous biometrics data to offer unprecedented user authentication and user identification capabilities. In this survey paper, we outline the nature of continuous authentication in IoT applications, highlight the key behavioral signals, and summarize the extant solutions from an AI perspective. Based on our systematic and comprehensive analysis, we discuss the challenges and promising future directions to guide the next generation of AI-based continuous authentication research.

Index Terms—Behavioral Biometric, Continuous Authentication, Data Mining, Internet of Things, Artificial Intelligence, Constrained Devices, Body Area Networks, Cyber-Physical Systems;

I. INTRODUCTION

With the flourishing of IoT, our daily life is being transformed by ambient intelligence [1] along with massive connected IoT devices ranging from smartphones and wearables to robots, autonomous vehicles and drones [2], [3]. The broad penetration of IoT devices in consumer market makes user authentication critically important to secure users have the appropriate right to access IoT devices [2] and to avoid the devastating damages caused by one attack occurring in the local vulnerable spots [4]. Apart from the security concerns, user authentication is beneficial for passive and customized services when the user switching occurs. For example, for one autonomous car shared among family members, the driving habits among family members differ significantly. To assist the drivers, different assistance strategies can be applied based on user identities [3]. Thus, user authentication can protect crucial information against potential attacks and offer customized services for improved user experience.

Due to the importance of user authentication, researchers and industries are increasingly studying the development of sophisticated methods to verify and recognize user identities. As shown in Fig. 1, authentication systems can be divided into three categories: knowledge-based, physiological biometric based, and behavioral biometric based solutions [2], [5]. Knowledge-based authentication explicitly requests user to enter credentials such as password, personal identification number (PIN) and graphical PIN to confirm the identity of an individual. Physiological biometric based authentication uses biological traits (e.g. fingerprint, iris, and facial images) and employs the machine learning methods to discriminate user identities. Behavioral biometrics including walking gait, keystroke and touchscreen dynamics are used for user authentication as well. Authentication systems can be classified into two sub-categories: user authentication to detect whether the

Fig. 1. An overview of credentials for user authentication and identification and their applications
user is one unauthorized visitor or genuine user and user identification to recognize whom the current user is.

The essence of authentication systems is to build the mapping relationship between users and objectives. According to the object-user mapping relationship, authentication systems can be categorized as Fig. 2. Among them, one-to-one mapping aims to verify whether the user is the genuine user or impostor for one privately-owned device (such as mobile phones and laptops) or one mobile application. One-to-many mapping provides the appropriate access control among multiple users for one object shared within a group of persons. In IoT systems, numerous smart devices are connected to provide pervasive services for one user (such as smart home and vehicle-to-vehicle systems [6]). In the dynamic environment, participants need to finish one session across shared IoT devices where complex and robust authentication schemes are needed [7]. The many-to-one mapping and many-to-many mapping fit well for the user authentication in complex dynamic environment.

Although numerous user authentication and identification methods are proposed, prior methods have several key drawbacks as it pertains to their fit with IoT applications:

**Vulnerability:** Prior systems are prone to a diverse range of attacks. For knowledge-based authentication, imposters can capture inputs by shoulder surfing and recording attacks [8]–[10], thermal attack [11], [12] and smudge attacks [13], [14]. For facial recognition, an adversary could conquer the facial detection through legitimate users’ facial photos. The fingerprint can be conquered by smudge attack [13]–[15] and forged by deep learning methods [16]. The automated speaker verification based on the personal characteristics of voices is subject to replay attacks [17], [18].

**Discreteness:** In general, user identification and authentication is executed once at the beginning of a session. If the authentication information is stolen or compromised, imposters can control the hacked accounts or IoT devices, resulting in devastating damages consequentially. In addition, the one-time user authentication is insecure in some scenarios. For example, in the ride-sharing platforms, registered drivers may subcontract the ride-assignments or share their registration to an unauthorized person, which could be dangerous for the riders [19]. Thus, the one-time authentication method is insecure and cannot provide seamless protection.

**Obtrusiveness:** Existing solutions require explicit inputs or actions, which are obtrusive for users by requiring extra user attention. They also cause a distraction from the ongoing tasks [20], [21]. For example, iris and facial recognition require users to stare at the camera in specific angles, which is unnatural and uncomfortable for users.

In recent years, the rapid proliferation of IoT devices such as smartphones, wearable devices and facility cameras has made it possible to seamlessly sense and track user behaviors. The analysis and mining of behavior fingerprints offer new opportunities for continuous authentication [22], [23]. In this paper, we provide a systematic overview about the continuous authentication based on behavioral biometrics from the perspective of AI. Our contributions in this paper are as follows:

- We provide a systematic overview of the key components and differentiators between user authentication and identification;
- We summarize the key elements of behavioral biometrics;
- We provide a summary of the emerging types of sensing technologies being integrated into emerging IoT technologies, with a specific focus on how the data they generate and common representations of these data;
- We present a general framework on how future researchers can develop innovative AI-based approaches for continuous user authentication and identification;
- We summarize emerging directions for future AI-based research in the aforementioned areas.

The remainder of this paper is organized as follows. In Section II, we characterize the nature of behavioral biometrics. In Section III, we propose a general framework for continuous user authentication from sensing and computing perspectives. Sections IV and V provide one systematic survey about data sensing and inference methods respectively. Finally, Section VI presents the open issues and challenges in CA based on behavioral biometrics and Section VII concludes this paper.

**II. Characterizing Behavioral Biometrics**

Behavioral biometrics refer to the unique behavioral traits that can be used for human authentication. Unlike the knowledge-based credentials and physiological biometrics shown in Fig. 1, behavioral biometrics identify people by how a user conducts the specified activity rather than by static information or physical characteristics. User authentication based on behavioral biometrics is characterized as **secure, continuous, transparent, and cost effective.**

**Secure:** In contrast to knowledge-based credentials and physiological biometrics, behavioral biometrics provide a dynamic modality that is completely passive and works in the background, making it impossible to copy or steal. Behavioral biometric data are extracted when users are performing one specified activities. Unlike the static authentication information, the nature of behavioral biometric data ensures that they cannot be forgotten, exchanged, and stolen. Moreover, the dynamics of activities makes it very difficult to forge behavioral biometrics. Authentication systems based on knowledge-based credentials and physiological biometrics are vulnerable to a variety of cyberattacks including shoulder surfing attacks,
smudge attacks, replay attacks, thermal attacks and adversary attacks [8]–[10], [13], [14]. Systems based on behavioral biometrics are secure and robust to the aforementioned cyberattacks.

**Continuous:** In the IoT era, user authentication is one crucial task to secure the connected devices, and it should not be a one-off event but rather a constant process. Unlike the one-time authentication that is enforced at the beginning of a session or login, continuous user authentication is an essential requirement to verify that users are who they claim to be on an ongoing basis. In order to achieve this goal, behavioral biometrics continuously profile a user’s behavior based upon the natural interactions without having to constantly interrupt users. The continuity of behavior makes it a nature way for continuous authentication with no distraction for users.

**Unobtrusive:** Unobtrusive sensing aims to monitor physical activities and behaviors continuously via sensors embedded in ambient environment or wearable sensors [24], and maximize the user experience to avoid disturbing users from the undergoing tasks [25]. Behavioral data can be sampled when users are interacting with IoT devices or ambient environments with no explicit input. Moreover, user authentication can be performed in a transparent and unobtrusive way with no distraction for users [26]. Previous attempts to continuously authenticate may have been too disruptive (e.g., prompts mid-session), but now by using unobtrusive sensing techniques users can be continuously authenticated without interruption. This feature is beneficial for the enhancement of user experience, and provides more secure protection for IoT devices.

**Cost Effective:** Physiological biometrics usually rely on customized hardware for information acquisition [27], [28]. This is often expensive in terms of costs and impedes the widespread adoption of physiological biometrics for user authentication. In contrast, behavioral biometrics can be observed and sampled with embedded sensors in IoT devices (e.g., microphone, touchscreen, accelerometer in smart phone and wearable devices) or public facilities (e.g., WiFi access point and surveillance camera). The widespread availability of IoT devices makes it possible to sense behaviors without extra hardware, which improves the acceptance of behavioral biometrics with low cost and ease of use.

III. OVERVIEW OF CONTINUOUS AUTHENTICATION

To guide the readers to understand the core concepts in this survey paper, we provide an overview of continuous authentication systems to illustrate what components should be included in one behavior-based continuous authentication system. Specifically, we present an abstract framework of continuous authentication and identification based on behavioral signals to highlight the primary components for user authentication. As shown in Fig. 3, it consists of four layers: behavioral signals, unobtrusive sensing, continuous computing and applications.

**Behavioral signals** are the collection of distinctive behavioral patterns or traits that can be used by one decision-making system to decide individual’s identity. A large number of pilot studies show that dynamics of keystroke, walking gaits, eye movements, touchscreen dynamics are suitable for continuous authentication.

**Unobtrusive sensing** summarizes the available sensors and feasible sensing strategies to capture the behavioral signals. The sensing modalities for behavioral signals are diverse. In the IoT era, the unprecedented sensing capability brings the opportunities to sense the behavior in different granularities with diverse sensors. For example, walking gaits can be captured by facility cameras, accelerometers built-in wearables, and WiFi signals that bounce off the walking individuals. IoT devices such as smartphones and wearables are infused into our daily life and can provide transparent, unobtrusive and continuous behavior sensing without additional attentions and actions required.

**Continuous computing** highlights the workflow for continuous authentication. The goals of CA based on behavioral biometrics are to detect whether the user has the right to access the IoT device or not (authentication), and to recognize whom the current user is (identification). Accordingly, continuous authentication based on behavioral biometrics can be divided into two categories: anomaly detection and classification. Anomaly detection methods can determine the abnormal patterns from the regular ones. For user identification, one predictive model is trained to maximize the inter-class differences (i.e., legitimate users vs. outliers).

**Applications** are typical scenarios where continuous authentication based on behavioral biometrics are applied. According to the property of devices, user authentication can be applied in three categories of IoT devices or scenarios including private devices, shared devices, and open application environments.

Apart from IoT device security issues for access control, behavioral biometrics can be applied for customized services in smart space. For example, for one autonomous car shared among family members, insurance companies can design tailored insurance policies according to individual’s driving patterns, where the continuous user identification based on driving patterns is the premise. For smart spaces, understanding the presence of users in the buildings is significantly important for providing more responsive and customized services [29].

---

Fig. 3. An abstract framework of continuous authentication and identification based on behavioral signals
IV. SENSING OF BEHAVIORAL BIOMETRICS

Numerous behavioral traits have been explored for continuous authentication. In this section, we analyze the commonly used behavioral traits for use authentication, and conduct comprehensive comparison from different dimensions including vulnerability, discreetness, obtrusiveness, and privacy.

A. Keystroke Dynamics

Keystroke dynamics characterize the typing rhythm such as keystroke length, distance between consecutive strokes, the pressure exerted on each key when the individual types characters, and others. To date, keystroke-powered authentication has been broadly explored for devices equipped with physical keyboards. With the emergence of touchscreens, when users enter characters via touchscreens, subtle changes of built-in sensors including accelerometer and gyroscope occur. Jointly combined with the status of built-in sensors, the keystroke timing, touch-typing, and keystroke pressure are distinctive features for user identification [30], [31]. Similarly, mouse usage dynamics have also been shown to serve as potential authentication cues [32], [33].

The advantages of analyzing keystroke dynamics include the unobtrusive data collection and continuous monitoring of typing behaviors when users interact with devices simultaneously. However, the keystroke dynamics vary in different scenarios such as walking, holding at hand, and putting on table [34]. As a result, keystroke-based user authentication is scenario-dependent, which requires the understanding of user scenarios and build the appropriate algorithms accordingly.

B. Touchscreen Dynamics

With the prevalence of touchscreen in IoT devices, sophisticated interaction patterns including pressure intensity and sliding dynamics when users interact with touchscreen enable the detection of user identification in an unobtrusive way [35]. One type of studies combine the touch patterns with conventional authentication method such as PIN codes or shaped-based drawing when individuals are running the log-in session. Even though more patterns are extracted to protect the devices against potential attacks, the authentication method is still static. For continuous authentication, Sitová et al. analyzed the micro-movement and orientation dynamics resulting from how a user grasps, holds and taps on the smart phone and leveraged the context and touchscreen dynamics for user authentication [23].

The authentication based on touch operations provides one natural way to collect user interaction data. Moreover, it makes the continuous user identification possible with no extra sensors and low computational load. However, touch operations vary among applications. Therefore, systematically studying application dependent touch patterns can help protect against unauthorized access of crucial mobile applications. This is especially true when individuals are likely to possess more than one mobile devices. When they interact with different devices, whether the touch dynamics are identical and can be transferred among different devices are still open questions.

C. Eye Movement

Driven by the internal interaction relationship between muscles and brain neural, eye movements including gaze and blinking are significantly different for individuals and are difficult to be mimicked and duplicated. Authentication based on eye movements can be divided into two categories in terms of data signals. Bioelectrical signals caused by eye movements and blinks are studied, and found that the accompanying electro-oculogram signals extracted from eye blinking were unique and rational as the biometrics for identification recognition tasks [36]. The dynamics of eye movements including pupillary response to stimuli, pupil size, velocity, acceleration, and spatial/geometric features are recorded and analyzed from video. Several studies demonstrated those patterns were intrinsic and could be applied for user identification [37]–[39].

However, eye-movement based solutions often use expensive and specialized monitor-mounted gaze trackers. Such explicit authentication methods may cause vigilance of the imposter and cover the camera of the tracker. These authentication methods also have high energy consumption for the continuous video recording and realtime video analysis. Furthermore, eye-movement based user identification can be obtrusive to an individual’s privacy.

D. Walking Gait

Identifying and authenticating based on walking gaits is an emerging biometric technology which recognizes users’ identities by analyzing walking patterns [26]. Based on the strategies of data acquisition, the sensing strategies of gait signals can be grouped as: facility cameras, floor sensors and device-free sensing.

Fig. 4. User authentication and identification based on walking gaits via different sensors including cameras, wearables, smart floor, and device-free sensing.

Fig. 4. User authentication and identification based on walking gaits via different sensors including cameras, wearables, smart floor, and device-free sensing. Identifying and authenticating based on walking gaits is an emerging biometric technology which recognizes users’ identities by analyzing walking patterns [26]. Based on the strategies of data acquisition, the sensing strategies of gait signals can be grouped as: facility cameras, floor sensors and wearables. Fig. 4 illustrates the interplay and relationships among the three components.

Vision based solutions record an individual’s gait patterns when walking via facility cameras. Then background segmentation techniques are used to extract features from recorded images to verify user identities [40], [41]. However, the vision-based solutions are subject to environments including illumination and camera angle [40], [42]. Furthermore, the
high computation consumption and privacy concerns make vision-based solutions infeasible for continuous authentication.

For floor-sensor based solutions, dense press sensors are deployed under floor to track the pressure dynamics or acoustic patterns when walking on the floor [43]. Its advantages include high resolution in terms of performance and unobtrusiveness for user verification [44]. However, floor-sensor based solutions are ideal for CA for two reasons. First, they often have sophisticated system design and high costs. Second, they only work in the enclosed environment with limited users and do not work in the open space with low scalability. Taken together, these limitations often result in floor-based systems being difficult to deploy.

Wearable sensor based solutions rely on sensors attached on different spots of the body (such as waist, hip, pocket) to capture the accompanying signals when walking, therefore enabling continuous verification of user identity [45]–[49]. However, primary studies are conducted in the laboratories with cumbersome prototype systems and expensive customized devices. Recently, more studies focus on the user verification based on off-the-shelf devices (e.g., mobile phones) [50]–[52]. The main advantage of using a wearable accelerometer sensor for gait recognition is that it provides unobtrusive verification without requiring user explicit actions. Especially, the accelerometer sensor has characters of small volume, low cost and can be easily integrated into the hardware of wearable devices.

Lately, walking gait recognition based on WiFi, millimeter wave and radio frequency identification (RFID) is gaining attention [53]. These studies assume that the unique walking gaits and body shapes entail distinctive disturbances in signals that can be used for user verification [54]. Wang et al. provided one comprehensive survey about the device-free sensing based on WiFi signals [55]. Several user verification systems based on WiFi signals are available including WiFiID [56], WiWho [57], WiFiU [58], and FreeSense [59]. Among them, FreeSense [59] is an unobtrusive system for indoor human identification based on the disturbed WiFi channel state information (CSI) signals when individuals walk through the line-of-sight (LOS) path between the source and the receiver of WiFi signals. FreeSense captures the disturbed waveforms when the user is walking across the LOS path, and applies the discrete wavelet transform and principal component analysis to extract shape features. The performance of FreeSense declines from 94.5% to 75.5% while the number of participants increases from two to nine. Luo et al. used RFID for gait recognition by monitoring the interruptions to RFID signals when one target user is blocking the signals between transmitter and receivers [53].

However, CA based on walking gaits has several challenges. First, prior studies were mainly conducted in a controlled environment. The robustness should be further evaluated in the physical world. Second, a person’s walking gaits can be altered by many factors, i.e., drunkenness, aging, carrying a load, and shoe type [60]–[62]. The model trained with the dataset collected in one situation may introduce bias when it is applied into other situations [63]. Third, prior studies mainly focus on the single-person gait detection. However, vision-based, floor-based, and WiFi based solutions achieve sub-optimal performances for multi-person scenarios.

### E. Body Gesture

With the popularity of wearable sensors, body gestures especially hand gestures have been widely studied for user authentication. Majority of user authentication based on body gestures attempt to verify or recognize the user identity based on a specified gesture performed. Among them, Li et al. [67] found that a person’s head movement patterns are unique when stimulated by music beats, and implemented the Headbanger, an authentication system that can authenticate uses by sensing head movements when listening to music beats based on the built-in accelerometer in Google phones. In addition, hand gesture and in-air-handwriting are studied for user authentication as well. Matsuo et al. [66] designed one authentication system based on the acceleration signals during the arm sweep action. Lu et al. [69] proposed a multifactor user authentication framework using both the motion signal of a piece of in-air-handwriting and the geometry of hand skeleton captured by a depth camera. However, user authentication based on body gestures are obtrusive as the users are required to perform certain movements or actions.

### F. Others

Chewing renders the changes of muscle tension with accompanying chewing sounds. Zou et al. proposed a human authentication mechanism that utilized the sounds generated by dental occlusion (i.e., tooth click) to unlock the mobile devices [70]. The prototype system, BiLock, relies on the microphone in mobile devices to record the sounds of dental occlusion, and verifies whether the current user is legitimate or not. Although BiLock is easy to use and requires no extra sensing unit, its performance is sensitive to the scenario. Furthermore, BiLock requires users to put the mobile device 5cm to 15cm away from lips, which is obtrusive for users and does not support the transparent sensing. Similarly, Bodybeat leverages the non-speech body sounds caused by food intake, breath, laughter, and cough for user identification [71].

In addition, breathing is used for user authentication as well by characterizing the subtle vibration caused by respiratory. For example, BreathPrint employs deep-learning models to effectively expression the acoustic features caused by breathing for the user authentication on resource-constrained devices [72]. To provide unobtrusive continuous authentication, Liu et al. proposed a continuous user verification system based on unique human respiratory-biometric characteristics extracted from the off-the-shelf WiFi signals [73].

### G. Summary of Behavioral Biometrics

As shown in Table I, each behavioral biometric has its pros and cons, and no single biometric is expected to effectively meet all the needs of any scenarios and applications [74]. The advantages of most user authentication systems based on behavioral biometrics include non-obtrusiveness with no extra user attentions required and non-vulnerability against...
TABLE I
SUMMARY OF BEHAVIORAL SIGNALS FOR USER AUTHENTICATION

<table>
<thead>
<tr>
<th>Category</th>
<th>Signal</th>
<th>Description</th>
<th>Vulnerability</th>
<th>Discreteness</th>
<th>Obtrusiveness</th>
<th>Privacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keystroke Dynamics</td>
<td>typing [30]–[33]</td>
<td>typing characters via keyboard or clicking via touchscreen when entering information</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td>Touch Operation</td>
<td>screen touch [23], [35]</td>
<td>finger movements and screen pressure on touchscreen</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td>Eye Movement</td>
<td>biosignal [36]</td>
<td>bioelectronic signals caused by eye movements</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>temporal [37]–[39]</td>
<td>dynamics of eye movement</td>
<td>✓</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Walking Gait</td>
<td>camera [40], [41]</td>
<td>capture the walking styles from video</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>floor sensor [43], [44]</td>
<td>features caused by users walking on the pressure sensors embedded in floor</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>wearable [45]–[49]</td>
<td>motions patterns tracked by wearable sensors</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>WiFi [55], [56]</td>
<td>Perturbations caused by unique walk styles when walking through the WiFi spectrum field</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>millimeter wave [64], [65]</td>
<td>interrupted radar signals between transmit antennas and receivers</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>RFID [53]</td>
<td>Perturbations caused by unique walk styles when walking through the RFID spectrum field</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td>Body Gesture</td>
<td>arm swing [66]</td>
<td>motion signals generated when a fixed behavior is performed</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>head movement [67]</td>
<td>head motions measured by head-worn devices</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>mouse movement [68]</td>
<td>Doppler profiles of acoustic signals caused by users’ speaking mouths</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>handwriting [69]</td>
<td>motion signals generated by in-air-handwriting</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Sounds of Behavior</td>
<td>dental occlusion [70]</td>
<td>unique acoustic patterns of teeth click</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>body sounds [71]</td>
<td>acoustic patterns caused by food intake, breath, and laugh</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>breathing [72], [73]</td>
<td>acoustic features caused by breathing</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>✓</td>
</tr>
</tbody>
</table>

cyberattacks. However, the existing authentication systems do not take advantage of the nature of behavioral traits to support the continuous authentication and do not consider the issue of privacy protection.

V. AI-BASED SOLUTIONS FOR USER AUTHENTICATION

In this section, we summarize the AI-based methods that are employed to recognize user identities based on behavioral fingerprints. In CA, intelligent algorithms including machine learning and deep learning are capable of determining the access control of IoT devices by checking user identities. As shown in Fig. 5, the pipeline of AI-based methods for CA consists of following major components: data pre-processing, feature extraction and classification algorithms. We further describe each component in the following sub-sections.

A. Data Pre-Processing

Data pre-processing is a critical procedure to distill high quality data out of the raw data that are generally incomplete, noisy, inconsistent, and redundant. As the inputs of continuous human identification based on behavioral patterns are sequential data, data filtering and data segmentation are necessary to reduce the noisy data and align the inputs.

For data filtering, numerous filters are applied to the sequential data to remove data that can be repetitive, irrelevant or even sensitive [56], [70]. For example, butterworth filter is applied to WiFi CSI data to remove the high frequency noisy data [56], [57]. Similarly, BiLock uses a 6-order butterworth filter to remove the out-of-band interference of dental clicks, and employs the wavelet de-noising to improve the signal to noise ratio [70].

For the continuous authentication, data segmentation seeks the effective regions from sequential data. The rule-based solutions including fixed threshold or fixed-size window are used over sequential data for segmentation. In WiFi-ID [56], two frequency bands are used to separate WiFi signals impacted by walking gaits from wavelet domain. The fixed-size window is applied [75], [76] to segment the dynamic swipe behavior and video records of walking gaits into fixed-length
slices respectively. However, rule-based solutions are sensitive to inputs, and tightly rely on the prior knowledge. Dynamic segmentation strategies are introduced to split the stream data dynamically. For example, BehaveSense [77] recognizes four touchscreen operations to separate the effective samples from stream data, and utilizes the sequential patterns of touchscreen operations for user identification. In addition, dynamic time warping (DTW) is widely used to find out the cycle of behaviors [57], [59], [67].

B. Feature Extraction

Feature engineering is to extract features of value that can represent users’ behavior comprehensively from the training dataset. The extracted features depend on sensors and applications. Statistical features refer to the measurements of interpreting both quantitative and qualitative data with standard statistics such as the root mean square, mean, standard deviation, and variance. Statistical features characterize the overall patterns of the given samples from macro perspectives. Due to the dynamics of user behaviors over time, frequency domain patterns of behaviors are of significance for describing the dynamics of signals. To obtain frequency domain patterns, sampled data are transformed by spectrum analysis to learn the frequency-domain features. As shown in Table II, DeepAuth employs the mel-frequency cepstral coefficients (MFCCs) to characterize the dynamics of vocal signals.

Generally, hand-crafted feature engineering heavily relies on the knowledge of domain experts and is time consuming to construct one complete feature set. As a result, it is the bottleneck of classification oriented tasks. In addition, as not all hand-crafted features are preeminently contributive to the verification of user identity, feature selection is optional.

<table>
<thead>
<tr>
<th>Type</th>
<th>Work</th>
<th>Signal</th>
<th>Features</th>
<th>Algorithm</th>
<th># of Users</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HMOG [23]</td>
<td>accelerometer, gyroscope, and magnetometer readings when a user tapping on the screen</td>
<td>HMOG features (grasp resistance and grasp stability features); 11 touchscreen features; keystroke dynamics</td>
<td>scaled Manhattan, scaled Euclidian, One-class SVM</td>
<td>100</td>
<td>EER=7.16% for walking; EER=10.05% for sitting</td>
</tr>
<tr>
<td></td>
<td>Omar et al. [32]</td>
<td>mouse dynamics, short-term Memory, visual scan and detection capability</td>
<td>statistical features of mouse dynamics</td>
<td>a statistical classifier based on Weighted-Sum</td>
<td>274</td>
<td>EER=2.11%</td>
</tr>
<tr>
<td></td>
<td>BiLock [70]</td>
<td>chewing sounds</td>
<td>13-order Melfrequency Cepstral coefficients (MFCCs) from occlusion sounds</td>
<td>SVM with radial basis function (RBF) kernel</td>
<td>100</td>
<td>FAR=1.1%, FRR=5.5%</td>
</tr>
<tr>
<td></td>
<td>Headbangger [67]</td>
<td>head movement with external rhythmic stimuli</td>
<td></td>
<td>dynamic time warping (DTW) distance</td>
<td>95</td>
<td>Accuracy=95.57%, FAR=4.43%</td>
</tr>
<tr>
<td></td>
<td>GlassGuard [78]</td>
<td>touch behavior, voice features, sensor data</td>
<td>MFCCs, touchscreen dynamics</td>
<td>One-class SVM, Threshold Random Walking</td>
<td>32</td>
<td>Accuracy=93%, FAR=3%</td>
</tr>
<tr>
<td></td>
<td>DeepAuth [75]</td>
<td>Motion Sensors</td>
<td>frequency domain signals</td>
<td>long short-term memory</td>
<td>47</td>
<td>Accuracy=96.7%</td>
</tr>
<tr>
<td></td>
<td>Song et al. [79]</td>
<td>Keystroke dynamics</td>
<td>key hold time, key latency time, key duration time</td>
<td></td>
<td>51</td>
<td>Accuracy=98%</td>
</tr>
<tr>
<td></td>
<td>Hong et al. [80]</td>
<td>Wave gesture based on accelerometer</td>
<td>singular value decomposition (SVD)</td>
<td>One-class SVM</td>
<td>8</td>
<td>Accuracy=92.83%; FPR (false positive rate)= 3.67%</td>
</tr>
<tr>
<td></td>
<td>George et al. [38]</td>
<td>Eye movement</td>
<td>fixation features and saccade features</td>
<td>Radial Basis Function Network</td>
<td>153</td>
<td>EER=4.59%</td>
</tr>
<tr>
<td>Identification</td>
<td>WiFi-ID [56]</td>
<td>Wi-Fi signals of walking gaits</td>
<td>7 time domain features and 3 frequency domain features</td>
<td>Sparse Approximation Classification</td>
<td>20</td>
<td>average accuracy of 93% to 77% from a group of 2 to 6 people</td>
</tr>
<tr>
<td></td>
<td>WiWho [57]</td>
<td>Wi-Fi signals of walking gaits</td>
<td>time domain and frequency domain features</td>
<td>decision tree</td>
<td>20</td>
<td>average accuracy of 92% to 80% from a group of 2 to 6 people</td>
</tr>
<tr>
<td></td>
<td>Batchuluun et al. [76]</td>
<td>thermal images of walking gaits</td>
<td>---</td>
<td>convolutional neural network (CNN)</td>
<td>80</td>
<td>EER=0.77%, Accuracy=99.9%</td>
</tr>
<tr>
<td></td>
<td>Mondal et al. [81]</td>
<td>swipe gesture</td>
<td>action duration, coordinate, distance, movement variability, orientation, velocity</td>
<td>Artificial Neural Network (ANN), Counter-Propagation ANN</td>
<td>41</td>
<td>Accuracy=95.45%</td>
</tr>
<tr>
<td></td>
<td>Abo-Zahhad et al. [36]</td>
<td>eye blinking</td>
<td>amplitude, position, area, energy, slope, duration et al.</td>
<td>linear discriminant analysis</td>
<td>25</td>
<td>Accuracy=97.3%; EER=3.7%</td>
</tr>
</tbody>
</table>
to rebuild one subset of attributes with least data loss. To mitigate the behavioral variability of mouse dynamics, Cai et al. proposed a unified framework of employing dimensionality reduction methods to extract predominant characteristics from the original feature space for enhanced performance, and found that variability reduction in feature engineering could enhance the authentication mechanisms [82]. On the other hand, due to these drawbacks of feature engineering, many researchers in this field have turned to deep learning based methods. Unlike conventional machine learning that relies on feature engineering, deep learning approaches utilize the complex neural network architecture to learn the representation of behaviors [76].

C. Anomaly-based User Authentication

For the privately-owned IoT devices, a large amount of positive examples from the legitimate users are available, while negative examples from imposters are rare. Therefore, supervised classification algorithms do not fit well to train predictive models when few negative examples are available [83]. To secure the privately-owned devices, anomaly detection solutions are applied to check whether the current user is one authorized user or one imposter. Behavioral examples from imposters are referred as anomalies or outliers, and the behavioral examples from genuine users are normal. Identifying outliers or anomaly detection is referred as one-class classification. For the user authentication task, one-class classification algorithms including one-class support vector machines (SVM) [84] and isolated forest [85] are widely adopted for imbalanced datasets with severely skewed class distributions.

One-class SVM [84] is a semi-supervised classification algorithm, and aims to find a hyperplane to enclose the majority of positive examples from the origin with the maximum margin [86]. Given the training vectors \( x_i \in \mathbb{R}^p \), the problem is formulated as below [83]:

\[
\min_{\mathbf{w}, \xi, \rho} \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{1}{v} \sum_i \xi_i - \rho \\
\text{subject to } \mathbf{w}^T \phi(x_i) \geq \rho - \xi_i, \xi_i \geq 0
\]

where \( \mathbf{w} \) is the normal vector of the separating hyperplane, and \( \xi_i \) are slack variables. The parameter \( v \in (0, 1] \) controls the trade-off between \( \mathbf{w} \) and slack variables \( \xi_i \). When \( \mathbf{w} \) and \( \rho \) are solved by solving Eq. 1, the decision function \( f(x) = \text{sgn}(\sum_i \alpha_i k(x_i, x) - \rho) \) will be positive for majority examples, where \( k(x_i, x) \) is a kernel function [87]. In the context of behavior-based user authentication, the workflow of one-class SVM is shown in Fig. 6 to highlight how the outliers are separated from the origin. An outlier is any data instance that lies outside the support of the training data. The original high-dimensional data can be projected into one feature space via one kernel function, where the hyperplane \( \mathbf{w} \) separates the training data from the origin by a maximal margin \( \rho/\|\mathbf{w}\| \) (Fig. 6). Data mapped to the same side of the origin will be given a negative one-class SVM value, whereas those mapped to the side of the training data will have positive values [88].

Isolated forest (iForest) [85], [89] is one unsupervised algorithm for anomaly detection. Its main idea is based on the observation that anomalies are few in number and much different from the rest of the data [85]. Specifically, iForest constructs an ensemble of binary search trees (named iTrees) in which anomaly points are isolated closer to the root of the tree. Each node in iTree has either two children or a leaf node with no child. For one \( p \)-dimension sample \( x_i \in \mathbb{R}^p \) from dataset \( D = \{x_1, x_2, \ldots, x_i, \ldots, x_n\} \), one feature \( a_i \in [1, p] \) and its split value \( V' \) are randomly selected. According to feature \( V_{a_i,k} \) for each input data \( x_i \), the \( V_{a_i,k} \) which is less than the \( V' \) is classified into left children set, the rest is classified into right children set. This process is repeated for the instances of left and right children nodes until (i) the incoming data set \( D \) has only one record or all data in \( D \) have the same value; (ii) the tree reaches the height limit \( l \) [85], [89].

Isolated forest is applied for user authentication due to the following reasons. First, the feature values extracted corresponding to the anomalies in original data are few and different. Second, iForest works well when handling extremely large data size and high-dimensional problems and in situations where training set does not contain any anomalies. Finally, since iForest has linear time complexity, a fast anomaly detection on resource-constrained IoT devices is crucial to report unauthorized access immediately.

D. Classification Algorithms for User Identification

User identification aims to recognize whom the current user is, and further determine whether the current user has the legitimate right to access the IoT devices or applications. Formally, behavior-based user identification can be formulated as follows. Given a dataset \( \{(x_1, y_1), \ldots, (x_n, y_n)\} \), where \( x_i = [x_{i1}, \ldots, x_{ip}] \) is the \( m \)-dimensional feature vector of samples; \( y_i \in \mathbb{C} \) is the corresponding class of one specific user; \( \mathbb{C} \) refers to the set of classes. The goal of user identification task is to learn a mapping function that predicts the label information for one given behavior sequence with least biases. According to the adopted classification algorithms, the extant behavior-based user identification systems can be divided into two categories: conventional and deep learning-based solutions.

1) Conventional classification: As shown in Table II, numerous hand-crafted features such as statistical features of
mouse dynamics, keystroke dynamics of touch-typing timing, and even frequency domain signals are proposed and a variety of supervised classification algorithms including SVM [70], random forest [64], Na"ive Bayes [30], and artificial neural network [81] are employed to bridge the mapping between feature sets and labels. However, user identification based on conventional classification algorithms relies on feature engineering, which involves computing explicit features specified by experts, resulting in algorithms designed to detect specific indicators. Hand-crafting feature is time consuming, labor intensive, and not suitable for rapidly evolving domains. Moreover, supervised learning algorithms are sensitive for the training dataset. Usually, the features used in those approaches are based on samples made from existing dataset. Due to the intra-user variation of behavioral biometrics [90], these solutions are not robust when faced with the changes of user behaviors, and cannot adjust the parameters accordingly.

2) Deep Learning-based classification: To improve the performance of identification systems, deep learning solutions are gaining popularity. Deep learning-based solutions can be divided into three categories according to the types of neural networks: convolutional neural network (CNN), recurrent neural network (RNN), and generative adversarial network (GAN).

As shown in Fig. 7(a), a typical CNN consists of convolutional layer, pooling layer, fully-connected layer, and softmax layer. In the convolutional layer, a convolution is a linear operation that involves the multiplication of a set of weights (also referred as filter) with the input. The convolutional layer creates one feature map by applying the same filter on the input repetitively to summarize the presence of a specific type of features in the input. The pooling layer operates upon each feature map and provides an approach to down sampling feature maps by summarizing the presence of features in patches of the feature map. Two common pooling methods are average pooling and max pooling. For the fully-connected layer, all the neurons in this layer are connected to every activation unit of the next layer operates on a flattened input where each input is connected to all the neurons. For one classification task, a softmax layer follows the final fully-connected layer immediately to limit the output of the function into the range 0 to 1, which can be interpreted directly as a probability of multiple classes.

CNN is widely adopted in user authentication and identification systems to detect personal patterns from fingerprints [91] and eyes [92], [93]. CNN is also commonly applied to detect the liveness of biometrics against presentation attacks [94]–[99]. For the security of smart vehicles, Xun et al. designed one driver fingerprinting device for the continuous user authentication of automobiles. The driver fingerprinting device is deployed in automobiles to collect the real-time driving data from on-board diagnostic port, and uses the CNN model to extract driver behavioral characteristics from the driving data. Finally, the extracted driving features are fed to SVM for illegal driver detection [100]. To secure the smart home, Qin et al. extracted the time and frequency features of sensors via CNN for each time slot, and fed the deep representation to RNN for user identification [101]. Batchuluun et al. [76] applied CNN to identify human identity based on the walking gaits extracted from videos. Unlike gait recognition in the controlled environment, Zhou et al. studied the user identification based on the walking gait in the wild, and proposed a hybrid model of CNN and RNN to learn robust gait feature representation including space and time domains features [102].

\[ h_t = \tanh(W_h h_{t-1} + W_y x_t) \]

(2)

RNN is a typical neural network to handle with sequence data, and the output of the prior state is forwarded as the input to the current state. This process can be formulated as shown in Eq. 2, where \( x_t \) is the input at time \( t \); \( h_t \) and \( h_{t-1} \) are the current state and previous state, respectively; \( W_h \) and \( W_y \) are feed-forward and recurrent weight matrices, respectively. The output \( y_t \) at time \( t \) is produced by combining the hidden state \( h_t \) with the weight matrix \( W_y \). However, RNNs are subject to
gradient vanishing and exploding problems [103]. In addition, the training of RNN is time-consuming and energy intensive [104]. To address those problems, long short term memory (LSTM) is proposed. The architecture of one cell in LSTM is shown in Fig. 7(b). One LSTM cell consists of forget gate, input gate, and output gate. The forget layer, denoted as \( f_t \), controls whether the previous state \( h_{t-1} \) is forwarded to the current state \( h_t \), where \( \sigma \) is a sigmoid activation function. The input gate \( i_t \) determines to what extent new memory is added into the cells state, and an output gate \( o_t \) regulates how gates at the next step will be affected by the previous cell state \( h_{t-1} \) and current input \( x_t \).

\[
\begin{align*}
    f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)
    \\
i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)
    \\
c_t &= \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)
    \\
o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)
    \\
h_t &= o_t \cdot \tanh(c_t)
\end{align*}
\]

Numerous user authentication and identification studies based on behavioral biometrics show that RNN and its variants are promising to process the sequential behavior data with overwhelming performance. For example, Zhang et al. extracted discriminate features from walking gaits monitored by smartphones and LSTM-based model for user identification [105]. Due to the limitation of resource-constrained devices in memory and computation, Chauhan et al. proposed one efficient authentication system based on breathing acoustics. They introduced model compression solutions including weight quantization and fully connected layer factorization to reduce the complexity of LSTM, and found that compressed LSTM outperformed other baselines with smaller model size, lower inference time, and more accurate [106]. Luo et al. used RFID for gait recognition by monitoring the interruptions to RFID signals and introduced attention mechanism in LSTM for robust user authentication [53]. Amini et al. [75] proposed a LSTM-based authentication framework that leveraged a user’s behavior captured by motion sensors while shopping online to continuously re-authenticate the user, providing security without compromising usability. DeepAuth [107] uses the unique motion patterns when users entering passwords as behavioural biometrics, and learns the deep representation of motion patterns via RNN-based model. Extensive experiments show that DeepAuth performs well for the security of resource-constrained devices with in both authentication performance and cost [107].

GAN is an unsupervised algorithm to train two competitive neural networks via a cooperative zero-sum game framework [108]. The GAN model consists of two sub-models: a generator model to generate new examples and a discriminator model to classify whether generated examples are real data or generated examples. As shown in Fig. 7(c), the generator network \( G(z) \) takes an random input \( z \) with probability distribution \( p(z) \) and generates a sample of synthetical examples. The discriminator network \( D(x) \) takes input either the real examples \( x \) from \( p_{data}(x) \) or synthetical examples generated by \( G(z) \), and attempts to predict whether the input is real or generated. The adversary learning of \( G(z) \) and \( D(x) \) can be represented mathematically as

\[
\min_{G} \max_{D} \mathbb{E}_{x \sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z \sim p(z)}[\log(1 - D(G(z)))]
\]

[109]. GAN has been widely adopted by the adversary to generate high fidelity human biometrics including fingerprints [110], facial and vocal biometrics [16], [111] to bypass the authentication systems. For example, DeepMasterPrints [110] employed GAN to generate synthetic image-level fingerprints. In DeepMasterPrints, two generator networks are trained via the Wasserstein GAN algorithm based on fingerprints scanned with a capacitive sensor and a dataset of inked and rolled fingerprints. Experimental results show that DeepMasterPrints is able to generate fingerprints that can easily bypass the popular commercial fingerprint matching systems.

E. Evaluation

For the evaluation of a verification system, the false acceptance rate (FAR) and the false rejection rate (FRR) are two types errors. FAR refers the likelihood that an unauthorized user is mistakenly accepted as a legitimate user; While FRR indicates the probability that a legitimate user is incorrectly rejected as an imposter. Verification systems should avoid those two error types. To balance the two error types in a verification system, equal error rate (EER) is introduced to predetermine the threshold value where FAR is equal to FRR. EER is a commonly accepted overall measure of system performance. The lower the EER, the higher the accuracy of the verification system.

VI. OPPORTUNITIES AND CHALLENGES

The continuous authentication based behavioral traits is an emerging paradigm facing several challenges. In this section, we enumerate the opportunities and challenges from the AI-perspective.

A. Evolution of Behavioral Biometrics

The performance of continuous authentication based behavioral biometrics could be impacted by scenarios and applications. For instance, alcohol, mood and carrying a backpack may affect the person’s gaits [112], [113]. Pupil size and eye movement dynamics vary with individuals’ physical status such as stress or fatigue [114]. In addition, touch patterns are sensitive with screen size and target applications. Obviously, user behaviors can be impaired by many facets including mood, health, and alcohol. How to rule out the impacts of situations on user behaviors in the CA systems are crucial for the robustness of behavior-based authentication systems.

User behaviors also change over time. For example, Galbally et al. analyzed the effects of age and aging on fingerprints, and found that fingerprint quality decreased linearly with age for elders [115]. The touchscreen typing-patterns are dynamic and impacted by health status [116]. However, what kinds of behavioral patterns can be used as behavioral biometrics for CA has not been studied yet. In addition, the extant behavior-based systems are static and cannot adjust
eliminates the less important connections in one pre-trained
the matrix singular value decomposition [127]. Neural pruning
a higher-order tensor with a sequence of linear operations on
sor decomposition reduces model complexity by expressing
pruning [125], [128] and parameter sharing [129]–[131]. Ten-
neural networks include tensor decomposition [126], [127],
significant problem. Methods to reduce the complexity of deep
models on resource-constrained devices, the following two
[123]. To execute the deep learning-based
system overhead
on resource-constrained IoT devices by
lions of hyper-parameters [124], time-consuming training,
ning algorithms achieved the state-of-the-art performance, deep
ables are resource-constrained with limited memory, power
supply and computing capability [123]. Although deep learn-
ing algorithms achieved the state-of-the-art performance, deep
learning models are becoming extremely complex with mil-
ions of hyper-parameters [124], time-consuming training,
significant energy strains [104], [125], and do not work well
on resource-constrained IoT devices by debilitating levels of
system overhead [123]. To execute the deep learning-based
models on resource-constrained devices, the following two
types of solutions are proposed.
Model compression of deep learning has become a sig-
ificant problem. Methods to reduce the complexity of deep
neural networks include tensor decomposition [126], [127],
pruning [125], [128] and parameter sharing [129]–[131]. Ten-
sor decomposition reduces model complexity by expressing
a higher-order tensor with a sequence of linear operations on
the matrix singular value decomposition [127]. Neural pruning
eliminates the less important connections in one pre-trained
model to reduce the computational cost by compressing hyper-
parameters [132] or multi-objective optimization based on
accuracy, latency and energy [133]. Parameter sharing [134]
is mainly applied in convolutional layers to reduce the size
of parameters by the assumption that the input going to be
processed by the network is decomposable into a set of local
regions with the same nature and thus each of them can be
processed with the same set of transformations [131], [134].

Edge computing is shedding lights on mobile devices ori-
ented deep learning [135], [136]. Via edge intelligence, the
complex and energy-intensive deep neural networks can be
partitioned into tiny subtasks, and be distributively executed
on neighboring devices or edges [137]–[139]. For example,
Kang et al. designed a light-weight scheduler to automatically
balance the computational offload between mobile devices and
servers by partitioning the neural network layers [137]. Xu et
al. proposed DeepWear to optimize the energy consumption of
wearables by offloading the deep learning tasks among mobile
devices [139].

Although there are many pilot studies about deep learning
on resource-constrained IoT devices, they are evaluated in
laboratory environments. As a result, the performance is not
generalizable to other contexts (e.g., smart home) due to the
extreme heterogeneity of IoT environments [140]. In addition,
CA on resource-constrained IoT devices is very sensitive for
the latency. A novel framework is essential to understand the
trade-off among accuracy, critical latency, and efficiency [141].

D. Emerging Malicious Attacks

CA based on behavioral biometrics cannot be easily attacked
by a random attacker. However, the nature of CA based on
behavioral biometrics does not imply that the CA systems are
secure.

First, most authentication systems based behavioral biomet-
rics are prototypes evaluated in constrained laboratory envi-
ronment with limited participants. Comprehensive evaluations
with a large number of participants are needed to investigate
the performance of existing behavior-based authentication
systems when faced with the potential attacks.

Second, IoT systems are faced with numerous security threats on physical, protocol, communication, and application
layers [28]. For example, in the low power wireless network,
the energy depletion attack can drain the batteries of devices
rapidly by forcing sensors or actuators to execute energy-

intensive tasks. Consequently, the entire network could fail
due to battery exhaustion [142]. For the electrical vehicles,
batteries could be attacked (e.g., draining energy) to reduce
driving range and increase driving range anxiety. Kang et
al. proposed a battery authentication method based on the
user habits to identify users that share a vehicle [143]. In
addition, the thermal attacks rely on the heat transferred
from users to interactive devices, and exploits heat traces
in the wake of user interaction with devices to uncover the
entered credentials [11]. On the communication layer, the
heterogeneous communication protocols are subject to attacks
such as eavesdropping, sinkhole, hello flood, collision [144],
[145]. Voice assistants such as Google Assistant, Amazon
Alexa, Facebook Portal, and Apple Siri are vulnerable to signal injection attacks on microphones based on the photoacoustic effect across large distances and through glass windows 1.

Third, AI is widely applied by imposters to hack the authentication and identification systems. For example, AI has been maliciously used by the imposters to infer the password. Snoopy demonstrates that it is possible to infer password entered on mobile devices by monitoring both accelerometers and gyroscopes [146]. Snoopy may fool the potential users as a harmless app to continuously monitor the motion sensors when users are taping the passwords, and uses bidirectional RNN for most commonly used password inference and encoder-decoder architecture with RNN models for universal password interference [146]. Another prevailing example is the usage of AI by the adversary to reconstruct the biometrics (i.e., replay attack). Deep generative models (DGMs) such as GANs and variational autoencoder (VAE) have been widely adopted to generate high fidelity human biometrics including fingerprints [110], facial and vocal biometrics [16], [111] to bypass the authentication systems.

E. Fusion of Behavioral Biometrics

Since prior solutions are insufficient to effectively provide secure protection in a broad range of IoT scenarios, multi-factor authentication (MFA) provides multiple layers of security to protect IoT devices against potential attacks through the fusion of behavioral biometrics [2].

The early form of MFA integrated multiple authentication schemes sequentially. For example, Hu et al. proposed a secure data backup scheme by integrating password and biometrics to overcome the potential attacks [147]. Multi-view representation learning for user authentication has emerged as a viable approach to process such data. This paradigm of machine learning aims to fuse multiple views (i.e., feature sets) to improve the performance [148], [149]. Multi-view representation learning can be categorized into two groups: multi-modal methods and multi-view methods.

Multi-modal solutions extract features from heterogeneous biometrics to build one classifier based on the ensembled features. Kim et al. designed one multi-modal authentication system by fusing features obtained from face, teeth and voice modalities [150] to secure mobile devices. Crawford et al. utilized the keystroke dynamics and speaker verification to enhance the authentication performance on mobile devices with a 67% reduction of explicit authentication [151]. EchoPrint emits nearly inaudible sound signals from the earpiece speaker to illuminate the user’s face. The extracted acoustic features from the echoes are combined with visual facial landmarks from the frontal camera to authenticate the user [152]. [153] integrates physiological biometrics, behavioral traits and online activities including search, shopping, and web browsing for user authentication. Kumar et al. utilizes LSTM to model the motion sensors and adopts CNN to extract gait patterns from video respectively, and Gray wolf optimizer has been used to optimize the parameters during fusion [154]. VAuth collects the body vibrations of the user and matches it with the speech signal received by the voice assistant’s microphone. By fusing multi-modal data, VAuth shows robust performance against potential attacks including replay attacks, mangled voice attacks, or impersonation attacks [155]. To against the replay attacks, REVOLT exploits the spectral differences between original and replayed voice signals, and combines the breathing rate extracted from WiFi signal while speaking to detect the liveness [156].

Multi-view based solutions for user authentication mainly extracted fine-grained feature maps from multi-view images. For example, Li et al. employed the multi-view deep representation learning to recognize one million celebrities from their face images captured in the real world [157]. On the other hand, multi-view learning can be applied for user authentication to address the incompleteness of obtained biometrics by fusing multiple views [158].

In addition, contextual information can enhance the performance of user authentication. Hintze et al. introduced dynamic factors like day and time, and location together with multi-modal biometrics to adjust the authentication scheme accordingly [159], [160] presents one context-based biometric authentication model, which chooses the appropriate authentication method dynamically according to the interaction form.

F. Cross-device Continuous Authentication

Increasingly, a large number of users are becoming multi-device users by interacting with multiple smart devices [117], [161]. For instance, more than 70% of online users access internet across multiple devices. 90% use multiple screens sequentially to accomplish a task over time 2. The complexity of multi-device, multi-user interaction presents significant challenges for cross-device continuous authentication. Prior studies assume the one-to-one mapping between user and device, and mainly focus on the user authentication in the single-device scenario. However, the relationships between users and devices in multi-user, multi-device scenarios is many-to-many. Due to the heterogeneity of devices, transferring one pre-trained user identification model from source domain to target domain is rarely studied. There are significant differences in interaction modality among heterogeneous devices. For devices with similar interaction modalities, transfer learning is one promising solution [117]. In addition, the co-location information of devices is useful for cross-device authentication. Hintze et al. proposed one multi-modal and cross-device authentication system based on behavioral and physiological biometrics (e.g., gait, voice, face, and keystroke dynamics) to reduce the manual burden of user verification according to the context like location, time of day and nearby devices [161].

G. Privacy Concerns

Behavior-based authentication systems are subject to privacy concerns, especially when they are adopted for personalized services. To address the privacy concerns, many privacy protection solutions are proposed to secure the crucial information in different authentication systems [162].

1https://lightcommands.com/
For the standalone authentication systems, although users are able to control the standalone client, they are likely to subject to exposing data to unauthorized third parties [162]. To secure the sensitive data in standalone systems, privacy impact assessment and surveillance impact assessment should be enforced to ensure conformance with legal and regulatory requirements [162]. Moreover, the anonymity and encryption are promising to protect against data exposure in privately-owned devices. For the centralized authentication systems, they may suffer from data exposure in transit and adversary attacks in computation models by carefully crafted adversarial samples [163]. Blockchain keeps the sensitive data private such that others cannot trace and infer sensitive data stored in the block [164]. However, the centralized authentication systems may suffer from the bottle problem due to the limitation of a single centralized server [7].

For the distributed authentication systems, blockchain and federated learning [165], [166] show the great potential to provide privacy-preserving authentication in collaborative applications. Federated learning is a machine learning technique that trains an algorithm across multiple decentralized edge devices or servers holding local data samples, without exchanging their data samples. This approach contrasts with traditional centralized machine learning techniques where all data samples are uploaded to one server, as well as to more classical decentralized approaches which assume that local data samples are identically distributed [165]–[167]. The nature of federated learning not only can prevent the data sharing among devices, but also avoid the enormous communication costs. In addition, other privacy-preserving machine learning approaches such as multi-party computation, homomorphic encryption are getting more attraction recently. For intrusted participants, the multi-party computation (MPC) is able to calculate a joint function in a decentralized network on the premise of ensuring privacy and independence of input [168].

VII. CONCLUSION

In IoT era, user authentication and identification are critical to ensure the security of connected things and the customization of passive services. However, conventional identification methods suffer from several key drawbacks including discreteness, obtrusiveness, and vulnerability. In this article, we propose the continuous authentication based on behavioral biometrics, characterize the key features of CA based on user behaviors (e.g., invulnerability, continuity, unobtrusiveness, and convenience), and summarize the existing CA solutions from sensing and computing. Based on this taxonomy, we discuss the challenges and open issues from the perspective of AI.

ACKNOWLEDGMENT

This work is supported by the national key research and development program of China under grant No.: 2019YFB2102200, by the ministry of health of China under grant No.: 2017ZX10303401-002 and 2017YFC1200302, by the natural science foundation of China under grant No.: 61902320, 71472175, 71602184, 71621002, by national science foundation under grant No. CNS-1850362 and OAC-1917117, and by the fundamental research funds for the central universities under grant No.:31020180QD140.

REFERENCES


Yunjie Liang is an associate professor from Northwestern Polytechnical University, China. He received his Ph.D. degree in computer science from Northwestern Polytechnical University in 2016. During 2012-2017, he worked at the University of Arizona as visiting scholar and post-doctoral researcher respectively. His research interests include pervasive computing, social computing and intelligent system.

Sagar Samtani is an assistant professor and Grant Thornton scholar in the operations and decision technologies department in the Kelley School of Business at Indiana University. He received his Ph.D. in Management Information Systems in the Artificial Intelligence Lab at the University of Arizona in 2018. His research interests include cyber threat intelligence, AI for cybersecurity, Dark Web analytics, and interpretable deep learning.

Bin Guo is a professor from Northwestern Polytechnical University, China. He received his Ph.D. degree in computer science from Keio University, Japan in 2009 and then was a post-doc researcher at Institut TELECOM SudParis in France. His research interests include ubiquitous computing and mobile crowd sensing.

Zhiwen Yu is a professor from Northwestern Polytechnical University, China. He has worked as an Alexander Von Humboldt Fellow at Mannheim University, Germany from Nov. 2009 to Oct. 2010, a research fellow at Kyoto University, Japan from Feb. 2007 to Jan. 2009. His research interests cover ubiquitous computing and HCI.