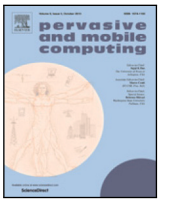
Pervasive and Mobile Computing 97 (2024) 101873

Contents lists available at ScienceDirect

Pervasive and Mobile Computing

journal homepage: www.elsevier.com/locate/pmc



SMARTCOPE: Smartphone Change Of Possession Evaluation for continuous authentication ☆

Nicholas Cariello [a], Seth Levine [a], Gang Zhou [b], Blair Hoplight [c], Paolo Gasti [a,*], Kiran S. Balagani [a]

[a] New York Institute of Technology, Old Westbury, NY, USA
[b] William & Mary, VA, USA
[c] Dominican University, Orangeburg, NY, USA

ARTICLE INFO

Keywords:
Continuous authentication
Change of possession
Mobile security
Machine learning
Activity recognition

ABSTRACT

The goal of continuous smartphone authentication is to detect when the adversary has gained possession of the user's device post-login. This is achieved by triggering re-authentication at fixed, frequent intervals. However, these intervals do not take into account external information that might indicate that the impostor has gained physical access to the user's device. Continuous smartphone authentication typically relies on behavioral cues, such as hand movement and touchscreen swipes, that can be collected without interrupting the user's activity. Because these behavioral signals are characterized by relatively high error rates compared to physiological biometrics, their use at fixed intervals leads to unnecessary interruptions to the user's activity in case of a false reject, *and* to not recognizing the impostor in case of a false accept.

To address these issues, in this paper we introduce a novel framework called SMARTCOPE: *Smartphone Change Of Possession Evaluation*. In this work, SMARTCOPE leverages smartphone movement signals collected during user activity to determine when the smartphone is no longer in the hands of its owner. When this occurs, SMARTCOPE triggers re-authentication. By using these signals, we are able to reduce the total number of re-authentication points while simultaneously lowering re-authentication error rates. Our analysis shows that our technique can reduce equal error rates by over 40%, from 7.8% to 4.6% using movement and keystroke features. Further, we show that SMARTCOPE can be used to transform a static (login-time) authentication system, such as face recognition, to a continuous re-authentication system, with a significant increase in security and limited impact on usability.

1. Introduction

Smartphones are commonly used to store and access private and sensitive information. However, over a quarter of Americans do not protect their smartphones using any authentication method [1], primarily because of the lack of usability of smartphone authentication techniques [2,3]. Users that choose to protect their devices rely on what is offered by their device's hardware and operating system, which typically includes PINs, passwords, fingerprint, and face recognition. These static authentication

☆ This project is supported by the National Science Foundation Secure and Trustworthy Cyberspace (SaTC) Grants No. 1619023/1618300 and 1814846, and New York Institute of Technology ISRC. The authors would like to acknowledge the effort of Kirthi Devleker, Diksha Chhabra, Maria Lombardo, Keyvan Chamani, Savitri Gadagi, Francheska Niveyro and Krutik Poojara for supporting the data collection effort associated with the paper.

* Corresponding author.
E-mail address: pgasti@nyit.edu (P. Gasti).

https://doi.org/10.1016/j.pmcj.2023.101873
Received 11 August 2023; Received in revised form 6 December 2023; Accepted 15 December 2023
Available online 23 December 2023

mechanisms are susceptible to guessing [4], spoofing [5], dictionary attacks [6] and side-channel attacks [7,8]. Further, these security mechanisms protect the device only at login, or "unlock" time. If the adversary gains physical access to an unlocked device, they can potentially access all the data stored on the device. The goal of continuous authentication is to mitigate this threat [3,9].

Continuous smartphone authentication typically uses behavioral signals that can be sampled continuously without interrupting the user, such as touchscreen interactions [10–13], gait signals [14,15], eye gazing [16], and hand movement, orientation, and grasp (HMOG) [17]. These signals can be leveraged to authenticate users post-login without requiring any additional effort from the user and thus, at least in principle, without impacting usability. Continuous authentication technology has also been successfully deployed in commercial settings, targeting primarily payments servicing industries [18]. For example, Sardine.ai [19] and BehavioSec [20] currently offer commercial products that include behavioral continuous authentication for smartphones. These systems have demonstrated reasonable accuracies in production settings [20].

However, continuous authentication has several important limitations. The identity of the user is verified at fixed intervals, irrespective of external cues that indicate that the smartphone is being continuously used by the same user since authentication. This impacts both security and usability. Security, because of high authentication latencies, due to authentication being triggered several seconds – or even minutes – after the adversary has grabbed the smartphone. Usability, because the relatively high false reject rates associated with behavioral signals can lead to users being prompted often to re-authenticate using more intrusive mechanisms such as passwords, PINs, or physiological biometrics.

In this paper, we address these two issues by introducing SMARTCOPE (Smartphone Change Of Possession Evaluation), a novel activity recognition framework that identifies whether the legitimate user has likely lost possession of their smartphone. When a "change of possession" event is detected, SMARTCOPE triggers re-authentication. This step can rely on either behavioral authentication mechanisms, or physiological biometrics. In the former case, the framework is able to improve both security and usability. False rejects are reduced by triggering behavioral re-authentication only when fairly confident that the device has changed hands, while false accept rates are reduced because the behavioral modality used to authenticate the user can be set to more stringent thresholds that make it more difficult for the adversary to evade detection. When instantiated with physiological biometrics, SMARTCOPE is the first system to transform a static authentication mechanism into a continuous re-authentication system. As a result, SMARTCOPE combined with physiological biometrics results in a drastic drop of post-unlock false accept rates: from 100% (no re-authentication is triggered when the attacker grabs the user's smartphone) to about 0.01% at 3.5% false reject rate as shown in our experiments.

Detection of change of possession is different from traditional authentication: rather than verifying the identity of the user, the goal is to identify cues that indicate interruptions in the user's possession of the phone. As a result, change of possession can be detected without user-specific training data, and therefore training can be performed offline on powerful computer systems, while prediction can be efficiently performed using energy-constrained devices such as smartphones.

Our experiments show that SMARTCOPE can significantly improve the authentication error rates of continuous authentication systems. For instance, our results show that SMARTCOPE can reduce the equal error rate of a continuous authentication system based on HMOG from 7.8% to 4.6%. While we use HMOG and face recognition to evaluate SMARTCOPE, this framework can be used with any biometric modality, including multi-modal biometrics.

In our evaluation, we also focus on the distinction between two types of "change of possession" events: (1) "give", where the user intentionally surrenders their device to another party; and "grab", where the device is forcefully taken from the hands of the user. To evaluate our approach, we collected and analyzed smartphone sensor data from 48 unique users.[1] This data was collected with the approval of New York Institute of Technology's Institutional Review Board (IRB). We show that our activity detection framework can be tuned to detect both types of events with high recall and precision (86.5% recall and 79.4% precision for grabs, and 74.6% recall and 83.3% precision for gives). This is important because, in practice, the user may want to be able to hand their device to another person without triggering re-authentication. For example, the user may want to give their device to a friend to show a picture. In this case, the device could be configured to lock highly-sensitive applications (e.g., banking apps) while allowing the use of less sensitive applications (e.g., a photo app). In contrast, when the device determines that it has just been grabbed from the hands of the user, it can lock itself and require explicit re-authentication.

Finally, we evaluate our framework against "rest" events. In this case, the user intentionally places the device on a table, thus relinquishing possession of it. We show that our framework can identify "rest" events with a recall of 74.9% and a precision of 86.5%.

1.1. Organization

The rest of this paper is organized as follows. Related research is reviewed in Section 2. In Section 3, we present SMARTCOPE and our methodology for evaluation, dataset, and parameters examined. Section 4 details our experiments, while Section 5 reports on our results. In Section 6 we discuss issues related to SMARTCOPE. We conclude in Section 7.

[1] As a further contribution of this work, we made the resulting dataset available for download: https://www.nyit-lamp.com/dataset/dataset3/

2. Related work

There is a substantial body of work that addresses continuous authentication using behavioral features, including movement [17,21,22], gait [23,24], pose [25], device interaction such as touch and swipe [26–29], and a combination of all of the above [30,31]. Similarly, human activity recognition (HAR) is a well-studied area due to its importance in human–computer interaction and mobile computing. The goal of HAR is to extract high-level knowledge about human activity from raw sensor data [32]. Research in the field began by using external or wearable sensors to acquire data from accelerometers and classify activities [33], and then shifted towards sensors onboard mobile devices and smartphones as they became available. The traditional activity recognition chain consists of data preprocessing, segmentation, feature extraction, and classification [34]. HAR has developed into its own field in machine learning [35] which has many applications in health care, fitness [36], human–machine interfacing and security. Extensive research has been completed using the inertial sensors (accelerometers and gyroscopes) to identify user activity [37–39]. A public domain dataset for HAR using smartphones has been released by Anguita et al. [40], and thoroughly analyzed in [37,41–43].

Smartphones can recognize many simple, daily activities with high accuracy. Common, repetitive behaviors such as walking, jogging, and climbing stairs have been recognized with accuracies ranging from 90% to 98% [37,44,45]. However, while it is straightforward to recognize repetitive activities that span long timescales, there is no established way of classifying short, non-periodic events. One study that tried to identify complex activities such as cleaning, cooking, taking medication, sweeping and washing hands using time-domain features extracted from accelerometers achieved accuracies as low as 50% [38].

Even if HAR techniques are successfully used for one recognition problem, they may not be as well adapted for a new problem domain [34,46]. Given the unknown duration of activities of interest, no single window length can be a "perfect fit" for all activities of interest [47]. Current research does not show a proven method to identify short, non-cyclic events such as giving or grabbing a smartphone.

Wójtowicz et al. [48] explores the idea of context-driven biometric authentication. Specifically, they look for a series of environmental cues that indicate which biometric modalities are more appropriate given the current context. For instance, their technique would not trigger voice authentication in a very loud environment.

Ramakrishnan et al. [49] demonstrates that the behavioral patterns of a user can be used to determine if they are still in control of the device. Their proposed solution works by detecting behavioral anomalies in addition to other data sources such as device location, activity, and application usage.

Chen et al. [50] introduces "device sharing awareness" (DSA), a system that detects when a user shares their device with a friend, and locks certain sensitive applications. In contrast to our work, DSA does not rely on a proper biometric authentication mechanism to determine whether the user is still in control/possession of her device. Further, DSA only considers scenarios in which the user willfully hands the phone to a friend. However, in our analysis, we assume that change of possession events correspond to attacks. Further, Chen et al. do not consider the detection of grab events, which is a key contribution of our work.

Liu et al. [51] presents a system to detect pickpocketing and grab-and-run theft based on accelerometer data. Liu et al. collected baseline data from 53 participants to serve as a baseline negative, after which simulated thefts were done in a lab setting with researchers. In their analysis, Liu et al. focus exclusively on the activity detection aspect of recognizing grab events. In contrast, we focus on both grab and give events, and combine this detection with behavioral and physiological biometrics. Our work demonstrates that, by combining change of possession detection with behavioral biometric authentication, we can achieve significant reductions in authentication error rates.

Riva et al. [52] aims to reduce the authentication error rates and energy footprint of active authentication by determining when not to re-authenticate based on user's activity. Their approach makes re-authentication decisions based on events such as: "is the user touching the screen?" These events include a large number of possibilities that vary across individuals depending on their app usage patterns and preferences.

We believe that SMARTCOPE and Riva et al. represent two different and potentially complementary approaches to address the same challenge: *when to re-authenticate*. However, the primary differences between these two approaches are related to: (1) the impact of false positives and false negatives on security and usability; (2) the number and variety of events that the two approaches must accurately identify; and (3) the availability of signals needed by SMARTCOPE and Riva et al.

With respect to (1), with SMARTCOPE a false positive has a potential impact on usability, because authentication is triggered when not needed. With Riva et al. a false positive impacts security because it implies that an adversarial event has been missed. The reverse is true for false negatives. As a result, the two approaches provide alternatives for systems with different security/usability tradeoffs and their tolerance to false positives/false negatives.

With respect to (2), SMARTCOPE deals with a small number of events which are relatively similar (e.g., measurable using the same sensors), and yet distinguishable. This simplifies the recognition task. On the other hand, Riva et al. must identify a larger set of events (Riva et al. lists nine activities), which potentially increases the complexity of the recognition task.

With respect to (3), Riva et al. requires signals that indicate continuity of a certain event. These signals might not be always available. In contrast, the signals leveraged by SMARTCOPE are triggered by adversarial events, and therefore they are always available during attacks.

3. SMARTCOPE rationale

The goal of SMARTCOPE is to detect when the user has lost possession of their smartphone. With SMARTCOPE, the smartphone consistently monitors for change of possession events. When the framework determines that a change of possession occurred, it triggers re-authentication. We model this behavior in terms of confidence on identity (CoI), which indicates the probability that the user is still in possession of the smartphone. Right after unlocking the device, CoI is high because the user just authenticated using a static authentication mechanism. As time passes, CoI decreases because the adversary may have gained possession of the smartphone. SMARTCOPE adds another dimension to this problem by indicating whether a change of possession event has likely occurred. When this happens, CoI drops to zero and re-authentication is triggered.

SMARTCOPE also addresses the tradeoff between security and usability defined by the choice of authentication window length. A longer authentication window typically allows more accurate authentication [17,53]. However, it also gives the adversary more time to use the system without being detected, thus increasing the impact of the attack. A shorter authentication window can be used to identify the adversary more quickly, but will lead to lower accuracy. SMARTCOPE aims to offer the best of both worlds. It provides high authentication accuracy and low authentication latency by triggering authentication only when it detects change of possession. Because events of interest are rare, the underlying biometric authentication mechanism can be skewed towards lower false accept rate (FAR), while SMARTCOPE overall does not incur in higher false reject rate (FRR), and thus can tolerate short authentication windows without impacting usability. Without SMARTCOPE, this would not be possible.

In order to demonstrate SMARTCOPE as a framework capable of integrating with any biometric, we use two distinct modalities: HMOG [17] (a behavioral biometric) and face [54] (a physiological biometric). Behavioral biometrics are used to authenticate the user transparently, i.e., without interrupting the user's activity. Physiological biometrics are used to authenticate the user by requiring the user to cooperate. Both are representative of biometrics that are currently deployed in modern smartphones.

Our analysis is based on the HMOG authentication error rates reported in Sitová et al. [17],[2] and on the CloudWalk MT 007 error rates under the "Visa Border" [54]. HMOG relies on information pertaining to hand movement, orientation, and grasp collected during several common smartphone activities. Sitová et al. [17] provides results for 120-s and 60-s authentication windows. In this work we use HMOG performance at 60-s windows because it provided good accuracy in the least amount of time. We also demonstrate that using 60-s HMOG windows with SMARTCOPE can outperform HMOG alone with a 120-s authentication windows.

3.1. Experiments setup

To demonstrate SMARTCOPE, we collected data from 48 unique subjects (15 male and 33 female) from a population of mostly graduate and undergraduate students. During data collection, the subject performed various typing activities, and experienced the following change of possession events: give, and grab. 26 subjects participated in two data collection sessions, while the remaining 22 subjects took part in one session. This resulted in a total of 74 sessions. Each session ranged from approximately 8 to 15 min with an average session duration of approximately 12 min. We used two Android smartphones (Google Pixel) to collect the raw sensor data from the accelerometer and gyroscope while users were typing answers to a series of questions and walking in a hallway. The smartphone sampled accelerometer and gyroscope signals on three axes at a rate of 100 Hz.

A series of events interrupted users during each data collection session. In each session, the subjects were asked three times to give the smartphone to the proctor in order to simulate a willful give event. Grabs events were simulated by abruptly taking the smartphone from the subject, without prior notice. We performed two grab events in each session. To collect data associated with rest (non-possession), the subjects were asked to play Jenga, which required them to place the smartphone on a nearby table without receiving any verbal cue. Give, grab, and rest make up the events of interest. Non-event indicates the user continuously possesses the phone. In our experiments, non-event signals primarily consist of the subject walking and typing answers to questions.

The phone front camera and a separate room camera were used to precisely identify the timing of events and to annotate the raw signals. In addition to the raw sensor data streams, we calculated acceleration and angular velocity magnitude as $mag = \sqrt{x^2 + y^2 + z^2}$. This resulted in a total of 8 signals: two magnitude signals, and six signals measured along the individual axes.

In order to segment and analyze sensor data, we applied a sliding window method. We derived statistical features (see Section 4) from the signals to create feature vectors. This technique consists of segmenting the data into fixed frame lengths with a certain amount of overlap. For example, with 2-s frames and 75% overlap, two consecutive frames are offset by 0.5 s and therefore share 75% of their data.

In human activity research, 50% is by far the most commonly used frame overlap (see for example [33]). However, the events examined in this study are shorter than other typical human activities, at approximately 1.5 to 4 s. As a result, both frame length and frame overlap had to be adjusted to account for the length of the events. Both frame length and frame overlap are parameters explored in the results section.

The ground truth of each frame was determined by finding the event that covers the majority of the frame. Due to the short nature of the events, a small frame size was necessary for all events to be recognized. From our observations, give and grab events tended to last between 1.39 s and 5 s. This affects the maximum frame size: if the frame size is set to 3 s, and the grab event length is 1.39 s (46.3% of the frame), then the frame would be labeled as non-event. As a result, the frame sizes examined in our experiments ranged from 0.5 to 2.5 s, with 0.5-s increments.

We also experimented with various frame overlaps, with percentages that ranged from no overlap to 90% overlap, in 10% increments. Frame length and overlap influence the total number of frames analyzed: as frame length decreases, the amount of frames increases; and as overlap increases, the number of frames increases.

[2] The HMOG dataset is available at https://hmog-dataset.github.io/hmog/

3.2. Evaluation metrics

The SMARTCOPE framework consists of two modules, (1) "change of possession" detection, and (2) biometric authentication. If a change of possession event is detected, a biometric re-authentication event is triggered. We evaluated our change of possession detection and the combined SMARTCOPE framework composed of change of possession detection and biometric authentication using distinct metrics.

We used recall, precision and F1-score as metrics to evaluate the classification performance of our change of possession detection. Correct classification can be described in terms of True Positives (TP) and True Negatives (TN). Incorrect classification can be described in terms of False Negatives (FN) and False Positives (FP). Recall (also known as sensitivity, or hit rate) is the percentage of correctly detected activities out of all actual instances of a particular class:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{1}$$

Given an actual class, recall determines if the classifier will predict it.

Precision (also known as positive predictive value) measures the likelihood that a detected instance of an activity corresponds to a real occurrence:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

Given a class prediction, precision determines how likely is it to be correct.

F-score combines the precision and recall rates into a single measure of performance. The F1-score weighs precision and recall equally and is defined as the harmonic mean of the two measures.

$$\text{F1} = \frac{2 \cdot \text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}} \tag{3}$$

The F1-score provides insight into the overall effectiveness of the classifier and is useful as a single measure to guide optimization.

We analyzed individual recall and precision scores for all events, with the primary optimizing metric being the grab F1-score. We also analyzed the influence of frame size and overlap value on recall, precision, and F1-score. The varying combinations of these parameters allowed for the resulting accuracy to be compared, thus allowing us to identify the ideal combination to be used in conjunction with a continuous authentication system.

Within the context of continuous authentication, we treat give and grab events as a "change of possession" event class (positive class), because both events indicate that the user is no longer in possession of the phone. Similarly, rest and non-events were jointly considered to be non-events (negative class).

With SMARTCOPE, a true positive is recorded when the change of possession detection module accurately identifies a change of possession event, *and* the biometric modality successfully identifies the impostor. As a result, the framework's TPR is $\text{TPR}_{\text{COP}} \times \text{TPR}_{\text{biometrics}}$, where TPR_{COP} indicates the true positive rate of the "change of possession" detection module, and $\text{TPR}_{\text{biometrics}}$ the true positive rate of the biometric authentication module.

We have a false positive only when the change of possession detection module incorrectly identifies a change of possession event, and the biometric module incorrectly identifies an impostor. As a result, the FPR of SMARTCOPE is calculated as $\text{FPR}_{\text{COP}} \times \text{FPR}_{\text{biometric}}$. In our experiments, we set the threshold for the change of possession detection to maximize TPR_{COP}. We then varied the threshold of the biometric component to determine overall TPR and FPR of the combined framework.

To measure the impact of using change of possession detection along with a biometric, we compared the equal error rate (EER) of SMARTCOPE with that of the underlying biometric modality.

4. "Change of possession detection" experiments

We extracted 24 features from accelerometer and 24 features from gyroscope time-domain signals. For each of the two sensors, the features consisted in mean, standard deviation (STD), median absolute deviation (MAD), minimum (Min), maximum (Max), and inter-quartile range (IQR) from the x, y, z, and magnitude components of accelerometer and gyroscope signals. Computation of these features require minimal resources. As a result, these features are well-suited for energy constrained mobile devices.

We used Random Forest classifier [55], a well-known ensemble method for pattern classification, because of its performance, speed, resilience against overfitting, ability to handle large feature matrices efficiently, and its ability to deal with unbalanced datasets. The classifier used is a multi-class classifier trained on four classes: "give", "grab", "rest", or "non-event". The parameters for Random Forest used in our experiments were 30 learning cycles and $n - 1$ maximum splits, where n is the number of frames in the training set.

We divided each session into fixed-length frames. Within each frame, data is viewed as a series of equal-length time intervals, and the dominant activity during that time is the label for that frame [56]. The evaluation determines whether the predicted label of a frame matches the corresponding ground truth. Consecutive frames are allowed to overlap. Depending on frame length and frame overlap, an event might occur in multiple frames.

Frame analysis is susceptible to multiple types of errors, including fragmentation, substitution, overfill, and under-fill [57]. Fragmentation errors occur when activity segments in the ground truth correspond to several segments in the recognition system output [34]. Substitution errors occur when a frame is incorrectly identified as another event but is temporally correct. Overfill and under-fill errors occur when the predicted labels match the ground truth, but the precise start and stop times do not align [56].

Table 1
Frame analysis of a "give" event showing an example of underfill that does not affect the detection of the event. In this example, frames 1 and 6, are correctly classified. Frames 2 to 5 have a ground truth of "give", but only frames 3 and 4 are identified correctly. Despite underfill, in practice our activity detection system would correctly determine that a "give" event happened within these frames.

Frame	Actual event	Predicted	Hit/Miss
1	Non-event	Non-event	Hit
2	Give	Non-event	Miss
3	Give	Give	Hit
4	Give	Give	Hit
5	Give	Non-event	Miss
6	Non-event	Non-event	Hit

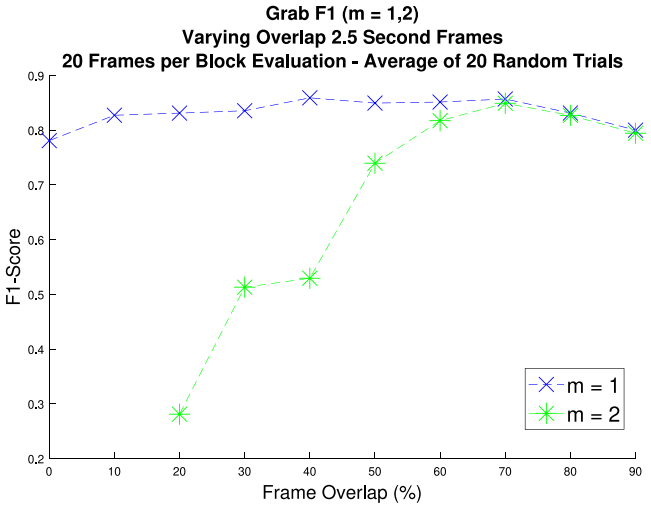


Fig. 1. Grab F1-scores with 2.5-s frames.

Table 1 shows an example of under-fill for the frame analysis of a give event. The first and last frame of the "give" event (frames 2 and 5) are labeled as "non-event", thus resulting in a boundary correspondence error.

In order to capture an event as it happens over time, we analyzed blocks of 20 consecutive frames. We set a threshold (m) for the number of frames within a block. If the number of frames exceeding this threshold in the block resulted in a positive event (give, grab), the entire block was considered as that event. With a threshold of 1, if any of the frames in the block were to be classified as either give, grab, or rest, the entire block would be labeled as the corresponding event. Looking at a block of frames allows the classification model to label events correctly despite small boundary correspondence issues. We determined that based on the F1-scores across multiple threshold values, the best threshold of frames to be set to $m = 1$. Fig. 1 shows how F1-scores vary for $m = 1$ and $m = 2$ with different frame overlaps.

In our analysis, when two or more frames within a block were classified as different events (e.g., the first as give, and the second as grab), we classified the entire block using the following hierarchy: grab > give > rest. This means that if any of the frames was classified as grab, the entire block would be classified as grab, and so on. This hierarchy was chosen because it is more important to correctly detect a grab event than a give event, and so on. This is due to the fact that, in general, the cost of a false negative for a grab event is higher than the cost of a false negative for a give event.

In order to increase generalization and validity of our results, we ran 20 trials. Each trial consisted of randomly assigning users to either training or testing data. For each trial, each combination of the parameters (frame size and frame overlap) were used to create a classification model. These parameters were chosen as they best represented the occurrence of a change of possession event. The classifier was trained 3 times using random forest algorithm with 5-fold cross validation on 75% of the data to create a frame-based predictor. A k value of 5 in k-fold cross validation uses 20% of the training data for validation and has been shown to produce results that do not exhibit excessively high bias or very high variance [58]. Due to the stochastic nature of a classification system that uses randomization, we used the average metrics of the three models to reduce variance. Each model is evaluated on the remaining 25% of the data (test set). For each combination of parameters, we report the average of recall, precision, and F1-score of 20 trials for grab, give, rest, and non-event.

Table 2
Experiment results with 2.5 s frames with 70% overlap.

Event	Precision (%)	Recall (%)	F1-score (%)
Grab	79.4	86.5	83
Give	83.3	74.6	78
Rest	86.5	74.9	79
Non-event	96	97.7	97

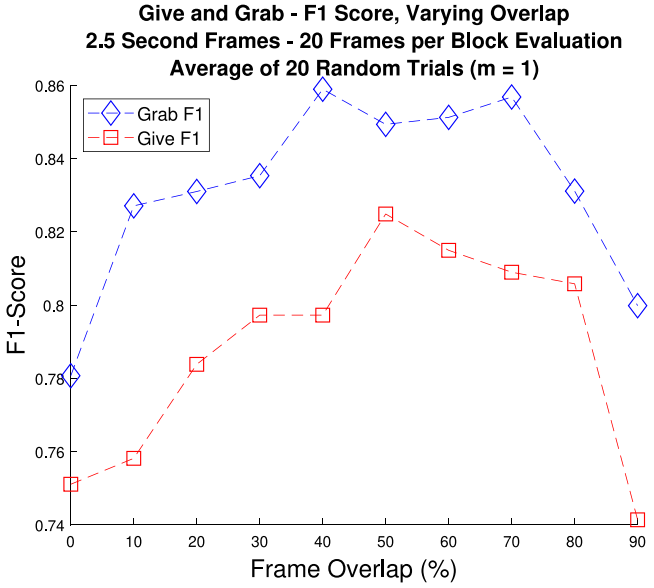


Fig. 2. Grab and give F1-scores at a frame length of 2.5 s, varying overlap.

5. Results

In this section we introduce the results of our experiments. We first discuss how frame size and overlap impacts change of possession detection accuracies (see Section 5.1). We then present authentication results for change of possession detection combined with the HMOG results [17] in Section 5.2. In Section 5.3 we present the results of change of possession detection results combined with CloudWalk MT 007 results [54]. The change of possession results are based on the dataset that we collected as part of this work (see Section 3.1).

5.1. Impact of frame size and overlap on change of possession detection

To evaluate our change of possession detection component, we analyzed combinations of frame lengths between 0.5 and 2.5 s, and overlaps between 0% and 90% with 10% increments. For frame size of 2.5 s, and overlap from 50% to 80%, we obtained F1-scores of at least 80% for all events examined. The average precision, recall, and F1-score of the 20 trials for each event at a frame size of 2.5 s and an overlap of 70% are shown in Table 2.

Some parameter combinations characterized by frame overlaps outside the 50%–80% range, and by frame sizes other than 2.5 s, led to either higher precision or recall for some of the events, or higher F1-score for non-events or give events. However, because our goal is to optimize for grab F1-score, we report parameters that maximize the results under this metric.

Generally, a higher overlap leads to higher F1-score. However, F1-scores for both give and grab events start to degrade once frame overlap reaches 80%. An overlap of 90% led to a significant drop in performance for all parameter configurations. Give and grab F1-scores with 2.5-s frame length are shown in Fig. 2, while Fig. 3 shows the relationship between frame length and F1-scores. In our experiments, the only frame length that consistently produced F1-scores above 0.8 for all events was 2.5 s.

To evaluate the performance of SMARTCOPE, we used 70% overlap with a 2.5-s window frame. We chose these parameters because they produced consistently good accuracies. Because the classifier used for change of possession is stochastic, we selected a representative instance as follows. From each training instance, we generated 60 ROC curves and calculated the corresponding areas under the curve (AUC). We then selected the instance with the median AUC. The resulting ROC is represented in Fig. 4.

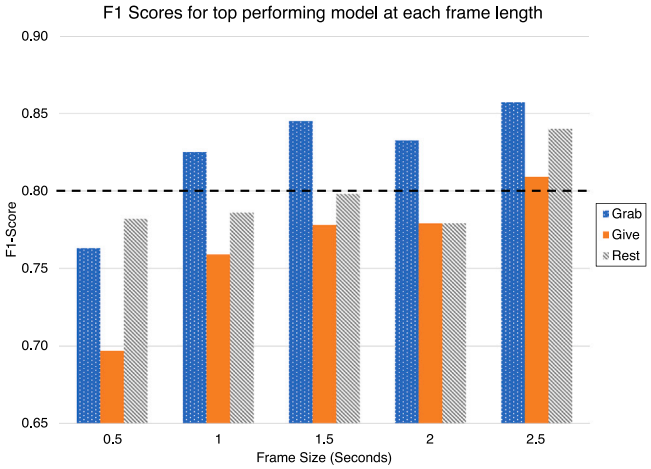


Fig. 3. F1-scores for top performing models at each frame length. A frame length of 2.5 s produces F1-scores above 0.8 for all activities of interest.

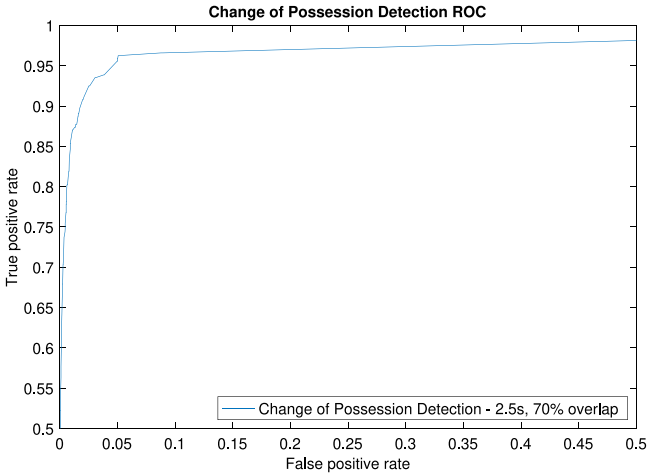


Fig. 4. The ROC curve plotted for the median trial for change of possession detection.

5.2. SMARTCOPE and HMOG

In this section we combine the results from the parameters selected in Section 5.1 parameters for change of possession detection with the results reported by Sitová et al. [17] with 60-s authentication intervals in the walking setting. Fig. 5 shows the resulting ROC curve, together with the ROC curve for HMOG. Our results show that SMARTCOPE leads to a meaningful improvement in overall accuracy. Specifically, EER improved from 7.8% to 4.6%—a decrease of over 40%. The advantage in performance of SMARTCOPE compared to the underlying biometric modality alone degrades as the TPR increases. The crossover point, however, is at around 31% FPR. This means that for all reasonable production settings, SMARTCOPE outperforms HMOG alone. Additionally, using SMARTCOPE with a 60-s HMOG window we were able to outperform HMOG alone with a 120 s window (7.16% EER).

5.3. SMARTCOPE and CloudWalk MT 007

We evaluated SMARTCOPE using the results for CloudWalk MT 007 face recognition with the "Visa Border" image set [54]. The goal, in this case, is not to improve the accuracy of this biometric modality—as with HMOG. Rather, the goal is to transform a static (login-time) biometric modality into a continuous-like or periodic authentication system. In the case of a static biometric, both TPR and FPR are 0% post-login, because the biometric modality is simply not triggered at periodic intervals within a session. In contrast, our results show that in the same setting SMARTCOPE is able to achieve an EER of 3.5%. That is, CloudWalk MT 007 used for login-time authentication has an EER of about 0.1% and when it is used for periodic authentication using SMARTCOPE is able to achieve EER of 3.5%. When using SMARTCOPE with CloudWalk MT 007, we were also able to achieve 3.55% false accept rate with 0.11% false reject rate.

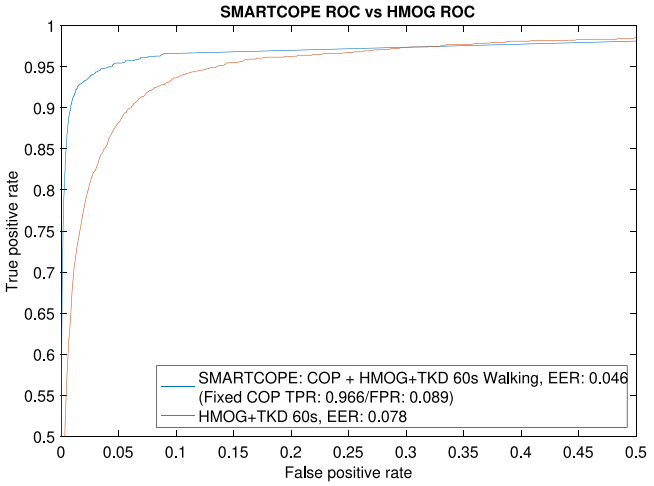


Fig. 5. The SMARTCOPE ROC curve plotted together with the ROC curve for HMOG and touch/keystroke dynamics (TKD) with 60 s authentication windows while walking (reproduced from [17]), along with the respective equal error rates.

6. Discussion and open problems

While our paper demonstrates the utility of SMARTCOPE as a proof of concept, further investigation is needed to address the following outstanding issues.

Identifying change of possession intent. While our work considers both "give" and "grab" events as adversarial, it is also possible to consider give as a benign event, and grab as a malicious event. This allows a fine-grained security response to a change of possession event. For instance, in case of a "give" event, SMARTCOPE could lock sensitive applications (e.g., a banking app) while leaving less-sensitive applications (e.g., a photo app) unlocked.

Another promising avenue is to use different biometrics modalities as a response to different change of possession events. For example, behavioral authentication could be used as a non-intrusive and passive biometric option when a "give" event is detected, while face recognition could be used in response to "grab" events. Further, depending on environment conditions (e.g., low light), alternative biometric modalities, such as fingerprint recognition, could be used to authenticate the user after a "grab" event.

Usability of SMARTCOPE. We believe that it would be meaningful for future work to address the usability of SMARTCOPE using a dataset collected in the field that incorporates information about users perception. This dataset could be used to compare the usability of SMARTCOPE with that of other context-driven continuous authentication systems under realistic operational conditions.

Recognition performance of rest events. The goal of this paper is to focus on identifying events that indicate adversarial change of possession. For this reason, we optimized our classifiers to detect "give" and "grab" events. As a result, the detection accuracy of "rest" events was impacted. We believe that this is a reasonable tradeoff, given the focus of our paper. However, for use cases where higher classification accuracies for "rest" events are needed, a standard thresholding approach over accelerometer signals can be used.

Harnessing the "rest" state. When a smartphone transitions from the resting state to an active state, SMARTCOPE could trigger authentication because the smartphone is unable to determine who picked it up. This approach could be instrumental in preventing unauthorized access when the device is momentarily left unattended.

7. Conclusion

In this paper we introduced a novel framework for continuous authentication based on the detection of change of possession. The latter represents a novel activity recognition problem that focuses on determining when a user lost possession of their device in the middle of a session, i.e., post-authentication.

Our results show two important findings. First, we demonstrated that combining change of possession with a behavioral biometric modality leads to a significant improvement in authentication accuracies. By combining HMOG [17] with change of possession, we were able to reduce the EER by over 40%, from 7.8 to 4.6%. Second, we showed that combining change of possession with a physiological biometric modality allows the resulting system to authenticate the user as soon as an adversarial event is detected, while at the same time having a negligible impact on the overall FNR. In particular, when combined with the CloudWalk MT 007 [54] face recognition model the resulting EER was 3.5%, which is remarkably low for a continuous authentication system.

We believe that this work is important because it provides a general framework for either improving the performance of behavioral biometrics, thus making them suitable for wider adoption, or for allowing re-authentication via physiological biometrics post-authentication.

CRediT authorship contribution statement

Nicholas Cariello: Conceptualization, Formal analysis, Investigation, Methodology, Visualization, Software implementation. **Seth Levine:** Conceptualization, Formal analysis, Investigation, Methodology, Validation, Visualization, Software implementation. **Gang Zhou:** Conceptualization, Formal analysis, Investigation, Methodology, Validation. **Blair Hoplight:** Conceptualization, Formal analysis, Investigation, Methodology, Supervision, Validation. **Paolo Gasti:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Supervision, Visualization. **Kiran S. Balagani:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Supervision, Visualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors have made the data public (link in paper).

References

[1] M. Anderson, Many smartphone owners don't take steps to secure their devices, 2017, URL http://www.pewresearch.org/fact-tank/2017/03/15/many-smartphone-owners-dont-take-steps-to-secure-their-devices/.

[2] M. Harbach, E. Von Zezschwitz, A. Fichtner, A. De Luca, M. Smith, It's a hard lock life: A field study of smartphone (un) locking behavior and risk perception, in: 10th Symposium on Usable Privacy and Security ({SOUPS} 2014), 2014, pp. 213–230.

[3] H. Khan, U. Hengartner, D. Vogel, Usability and security perceptions of implicit authentication: Convenient, secure, sometimes annoying, in: Eleventh Symposium on Usable Privacy and Security ({SOUPS} 2015), 2015, pp. 225–239.

[4] G. Ye, Z. Tang, D. Fang, X. Chen, K. In Kim, B. Taylor, Z. Wang, Cracking android pattern lock in five attempts, in: Network and Distributed System Security Symposium, 2017, http://dx.doi.org/10.14722/ndss.2017.23130.

[5] R. Gonzalo, et al., Attacking a smartphone biometric fingerprint system: A novice's approach, in: 2018 International Carnahan Conference on Security Technology (ICCST), 2018, http://dx.doi.org/10.1109/CCST.2018.8585726.

[6] P. Bontrager, J. Togelius, N. Memon, Deepmasterprint: Generating fingerprints for presentation attacks, 2017, arXiv preprint arXiv:1705.07386.

[7] D. Shukla, R. Kumar, A. Serwadda, V.V. Phoha, Beware, your hands reveal your secrets!, in: Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, CCS '14, ACM, New York, NY, USA, 2014, pp. 904–917, http://dx.doi.org/10.1145/2660267.2660360.

[8] D. Shukla, V.V. Phoha, Stealing passwords by observing hands movement, IEEE Trans. Inf. Forensics Secur. 14 (12) (2019) 3086–3101, http://dx.doi.org/10.1109/TIFS.2019.2911171.

[9] W. Lee, R.B. Lee, Sensor-based implicit authentication of smartphone users, in: 2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), 2017, pp. 309–320, http://dx.doi.org/10.1109/DSN.2017.21.

[10] M. Frank, R. Biedert, E. Ma, I. Martinovic, D. Song, Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication, IEEE Trans. Inf. Forensics Secur. 8 (1) (2013) 136–148, http://dx.doi.org/10.1109/TIFS.2012.2225048.

[11] X. Zhao, T. Feng, W. Shi, I.A. Kakadiaris, Mobile user authentication using statistical touch dynamics images, IEEE Trans. Inf. Forensics Secur. 9 (11) (2014) 1780–1789.

[12] N. Zheng, K. Bai, H. Huang, H. Wang, You are how you touch: User verification on smartphones via tapping behaviors, in: 2014 IEEE 22nd International Conference on Network Protocols, 2014, pp. 221–232, http://dx.doi.org/10.1109/ICNP.2014.43.

[13] B. Shrestha, M. Mohamed, S. Tamrakar, N. Saxena, Theft-resilient mobile wallets: Transparently authenticating NFC users with tapping gesture biometrics, in: Proceedings of the 32nd Annual Conference on Computer Security Applications, ACSAC '16, ACM, New York, NY, USA, 2016, pp. 265–276, http://dx.doi.org/10.1145/2991079.2991097, URL http://doi.acm.org/10.1145/2991079.2991097.

[14] M.O. Derawi, C. Nickel, P. Bours, C. Busch, Unobtrusive user-authentication on mobile phones using biometric gait recognition, in: 2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2010, pp. 306–311, http://dx.doi.org/10.1109/IIHMSP.2010.83.

[15] C. Nickel, T. Wirtl, C. Busch, Authentication of smartphone users based on the way they walk using k-nn algorithm, in: 2012 Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IEEE, 2012, pp. 16–20.

[16] J. Yin, J. Sun, J. Li, K. Liu, An effective gaze-based authentication method with the spatiotemporal feature of eye movement, Sensors 22 (8) (2022) http://dx.doi.org/10.3390/s22083002, URL https://www.mdpi.com/1424-8220/22/8/3002.

[17] Z. Sitová, J. Šeděnka, Q. Yang, G. Peng, G. Zhou, P. Gasti, K.S. Balagani, HMOG: New behavioral biometric features for continuous authentication of smartphone users, IEEE Trans. Inf. Forensics Secur. 11 (5) (2016) 877–892, http://dx.doi.org/10.1109/TIFS.2015.2506542.

[18] 2021 Fraud Transformation Survey: Detecting and Preventing Emerging Schemes. URL https://www.biocatch.com/resources/2021-fraud-transformation-survey-detecting-and-preventing-emerging-schemes.

[19] Z. Shaikh, How can behavioral biometrics prevent fraud? 2023, URL https://www.sardine.ai/blog/how-can-behavioral-biometrics-prevent-fraud.

[20] BehavioSec, Accuracy report for native mobile application, 2016, URL https://behaviosec.com/resources?resourceid=4369.

[21] M. Abuhamad, T. Abuhmed, D. Mohaisen, D. Nyang, Autosen: Deep-learning-based implicit continuous authentication using smartphone sensors, IEEE Internet Things J. 7 (6) (2020) 5008–5020, http://dx.doi.org/10.1109/JIOT.2020.2975779.

[22] J. Dybczak, P. Nawrocki, Continuous authentication on mobile devices using behavioral biometrics, in: 2022 22nd IEEE International Symposium on Cluster, Cloud and Internet Computing (CCGrid), 2022, pp. 1028–1035, http://dx.doi.org/10.1109/CCGrid54584.2022.00125.

[23] L. He, C. Ma, C. Tu, Y. Zhang, Gait2Vec: Continuous authentication of smartphone users based on gait behavior, in: 2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD), 2022, pp. 280–285, http://dx.doi.org/10.1109/CSCWD54268.2022.9776313.

[24] M. Abuhamad, A. Abusnaina, D. Nyang, D. Mohaisen, Sensor-based continuous authentication of smartphones' users using behavioral biometrics: A contemporary survey, IEEE Internet Things J. 8 (1) (2021) 65–84, http://dx.doi.org/10.1109/JIOT.2020.3020076.

[25] X. Huang, S. Nishimura, B. Wu, A pose detection based continuous authentication system design via gait feature analysis, in: 2022 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), 2022, pp. 1–5, http://dx.doi.org/10.1109/DASC/PiCom/CBDCom/Cy55231.2022.9927959.

[26] R. Murmuria, A. Stavrou, D. Barbará, D. Fleck, Continuous authentication on mobile devices using power consumption, touch gestures and physical movement of users, in: H. Bos, F. Monrose, G. Blanc (Eds.), Research in Attacks, Intrusions, and Defenses, Springer International Publishing, Cham, 2015, pp. 405–424.
[27] Z. Shen, S. Li, X. Zhao, J. Zou, MMAuth: A continuous authentication framework on smartphones using multiple modalities, IEEE Trans. Inf. Forensics Secur. 17 (2022) 1450–1465, http://dx.doi.org/10.1109/TIFS.2022.3160361.
[28] A.T. Kiyani, A. Lasebae, K. Ali, Continuous user authentication based on deep neural networks, in: 2020 International Conference on UK-China Emerging Technologies (UCET), 2020, pp. 1–4, http://dx.doi.org/10.1109/UCET51115.2020.9205446.
[29] Y. Barlas, O.E. Basar, Y. Akan, M. Isbilen, G.I. Alptekin, O.D. Incel, DAKOTA: Continuous authentication with behavioral biometrics in a mobile banking application, in: 2020 5th International Conference on Computer Science and Engineering (UBMK), 2020, pp. 1–6, http://dx.doi.org/10.1109/UBMK50275.2020.9219365.
[30] D. Hintze, M. Füller, S. Scholz, R.D. Findling, M. Muaaz, P. Kapfer, W. Nüßer, R. Mayrhofer, CORMORANT: On implementing risk-aware multi-modal biometric cross-device authentication for android, in: Proceedings of the 17th International Conference on Advances in Mobile Computing & Multimedia, in: MoMM2019, Association for Computing Machinery, New York, NY, USA, 2020, pp. 117–126, http://dx.doi.org/10.1145/3365921.3365923, URL https://doi-org.arktos.nyit.edu/10.1145/3365921.3365923.
[31] C. Wu, K. He, J. Chen, R. Du, Y. Xiang, CaIAuth: Context-aware implicit authentication when the screen is awake, IEEE Internet Things J. 7 (12) (2020) 11420–11430, http://dx.doi.org/10.1109/JIOT.2020.3006870.
[32] J. Wang, Y. Chen, S. Hao, X. Peng, L. Hu, Deep learning for sensor-based activity recognition: A survey, 2017, CoRR abs/1707.03502. arXiv:1707.03502. URL http://arxiv.org/abs/1707.03502.
[33] L. Bao, S.S. Intille, Activity recognition from user-annotated acceleration data, in: A. Ferscha, F. Mattern (Eds.), Pervasive Computing, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, pp. 1–17.
[34] A. Bulling, U. Blanke, B. Schiele, A tutorial on human activity recognition using body-worn inertial sensors, ACM Comput. Surv. 46 (3) (2014) 33.
[35] O.D. Lara, M.A. Labrador, A survey on human activity recognition using wearable sensors, IEEE Commun. Surv. Tutor. 15 (3) (2013) 1192–1209.
[36] A. Avci, S. Bosch, M. Marin-Perianu, R. Marin-Perianu, P. Havinga, Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey, in: 23th International Conference on Architecture of Computing Systems 2010, VDE, 2010, pp. 1–10.
[37] D. Anguita, A. Ghio, L. Oneto, X. Parra, J.L. Reyes-Ortiz, Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine, in: J. Bravo, R. Hervás, M. Rodríguez (Eds.), Ambient Assisted Living and Home Care, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 216–223.
[38] S. Dernbach, B. Das, N.C. Krishnan, B.L. Thomas, D.J. Cook, Simple and complex activity recognition through smart phones, in: 2012 Eighth International Conference on Intelligent Environments, 2012, pp. 214–221, http://dx.doi.org/10.1109/IE.2012.39.
[39] J.R. Kwapisz, G.M. Weiss, S.A. Moore, Activity recognition using cell phone accelerometers, SIGKDD Explor. Newsl. 12 (2) (2011) 74–82, http://dx.doi.org/10.1145/1964897.1964918, URL http://doi.acm.org/10.1145/1964897.1964918.
[40] D. Anguita, A. Ghio, L. Oneto, X. Parra, J.L. Reyes-Ortiz, A public domain dataset for human activity recognition using smartphones, in: ESANN, 2013.
[41] A. Ignatov, Real-time human activity recognition from accelerometer data using convolutional neural networks, Appl. Soft Comput. 62 (2018) 915–922.
[42] W. Jiang, Z. Yin, Human activity recognition using wearable sensors by deep convolutional neural networks, in: Proceedings of the 23rd ACM International Conference on Multimedia, ACM, 2015, pp. 1307–1310.
[43] R. San-Segundo, J.D. Echeverry-Correa, C. Salamea, J.M. Pardo, Human activity monitoring based on hidden Markov models using a smartphone, IEEE Instrum. Meas. Mag. 19 (6) (2016) 27–31.
[44] A. Reiss, G. Hendeby, D. Stricker, A competitive approach for human activity recognition on smartphones, in: European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2013), 24-26 April, Bruges, Belgium, ESANN, 2013, pp. 455–460.
[45] B. Romera-Paredes, M.S. Aung, N. Bianchi-Berthouze, A one-vs-one classifier ensemble with majority voting for activity recognition, in: ESANN, 2013.
[46] O. Banos, J.-M. Galvez, M. Damas, H. Pomares, I. Rojas, Window size impact in human activity recognition, Sensors 14 (4) (2014) 6474–6499.
[47] A. Reiss, D. Stricker, Creating and benchmarking a new dataset for physical activity monitoring, in: Proceedings of the 5th International Conference on Pervasive Technologies Related to Assistive Environments, ACM, 2012, p. 40.
[48] A. Wójtowicz, K. Joachimiak, Model for adaptable context-based biometric authentication for mobile devices, Pers. Ubiquitous Comput. 20 (2016) http://dx.doi.org/10.1007/s00779-016-0905-0.
[49] A. Ramakrishnan, J. Tombal, D. Preuveneers, Y. Berbers, PRISM: Policy-driven risk-based implicit locking for improving the security of mobile end-user devices, in: Proceedings of the 13th International Conference on Advances in Mobile Computing and Multimedia, in: MoMM 2015, Association for Computing Machinery, New York, NY, USA, 2015, pp. 365–374, http://dx.doi.org/10.1145/2837126.2837157.
[50] J. Chen, U. Hengartner, H. Khan, Sharing without scaring: Enabling smartphones to become aware of temporary sharing, in: Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022), USENIX Association, Boston, MA, 2022, pp. 671–685, URL https://www.usenix.org/conference/soups2022/presentation/chen.
[51] X. Liu, D. Wagner, S. Egelman, Detecting phone theft using machine learning, 2018, pp. 30–36, http://dx.doi.org/10.1145/3209914.3209923.
[52] O. Riva, C. Qin, K. Strauss, D. Lymberopoulos, Progressive authentication: Deciding when to authenticate on mobile phones, in: Proceedings of the 21st USENIX Conference on Security Symposium, Security '12, USENIX Association, Berkeley, CA, USA, 2012, p. 15, URL http://dl.acm.org/citation.cfm?id=2362793.2362808.
[53] L. Yang, Y. Guo, X. Ding, J. Han, Y. Liu, C. Wang, C. Hu, Unlocking smart phone through handwaving biometrics, IEEE Trans. Mob. Comput. 14 (5) (2015) 1044–1055, http://dx.doi.org/10.1109/TMC.2014.2341633.
[54] NIST cloudwalk MT 007 evaluation, 2023, URL https://pages.nist.gov/frvt/reportcards/11/cloudwalk_mt_007.html.
[55] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.
[56] D. Minnen, T. Westeyn, T. Starner, J. Ward, P. Lukowicz, Performance metrics and evaluation issues for continuous activity recognition, Perform. Metr. Intell. Syst. 4 (2006) 141–148.
[57] J.A. Ward, P. Lukowicz, H.W. Gellersen, Performance metrics for activity recognition, ACM Trans. Intell. Syst. Technol. 2 (1) (2011) 6.
[58] T. Hastie, G. James, R. Tibshirani, D. Witten, An introduction to statistical learning with applications in R, 2013.