When Your Fitness Tracker Betrays You: Quantifying the Predictability of Biometric Features Across Contexts

Simon Eberz*, Giulio Lovisotto*, Andrea Patanè*, Marta Kwiatkowska*, Vincent Lenders† and Ivan Martinovic*
*Department of Computer Science, University of Oxford, UK
Email: firstname.lastname@cs.ox.ac.uk
†armasuisse, Switzerland
Email: vincent.lenders@armasuisse.ch

Abstract—**Attacks on behavioral biometrics have become increasingly popular. Most research has been focused on presenting a previously obtained feature vector to the biometric sensor, often by the attacker training themselves to change their behavior to match that of the victim. However, obtaining the victim's biometric information may not be easy, especially when the user's template on the authentication device is adequately secured. As such, if the authentication device is inaccessible, the attacker may have to obtain data elsewhere.**

In this paper, we present an analytic framework that enables us to measure how easily features can be predicted based on data gathered in a different context (e.g., different sensor, performed task or environment). This framework is used to assess how resilient individual features or entire biometrics are against such cross-context attacks. In order to be able to compare existing biometrics with regard to this property, we perform a user study to gather biometric data from 30 participants and five biometrics (ECG, eye movements, mouse movements, touchscreen dynamics and gait) in a variety of contexts. We make this dataset publicly available online.

Our results show that many attack scenarios are viable in practice as features are easily predicted from a variety of contexts. All biometrics include features that are particularly predictable (e.g., amplitude features for ECG or curvature for mouse movements). Overall, we observe that cross-context attacks on eye movements, mouse movements and touchscreen inputs are comparatively easy while ECG and gait exhibit much more chaotic cross-context changes.

1. Introduction

Biometric authentication is a popular approach to address the shortcomings of passwords (e.g., bad memorability and password reuse). The most common approaches are fingerprint scanning and face recognition, both of which are used in scenarios ranging from smartphone security to border controls. However, both can easily be observed and replicated by an attacker, resulting in a security vulnerability. Fingerprints are easily lifted off smooth surfaces (such as coffee mugs) or captured through high-resolution photographs. 2D photos of a victim's face are readily available through social media profiles. In addition, fingerprints in particular require specialized sensors.

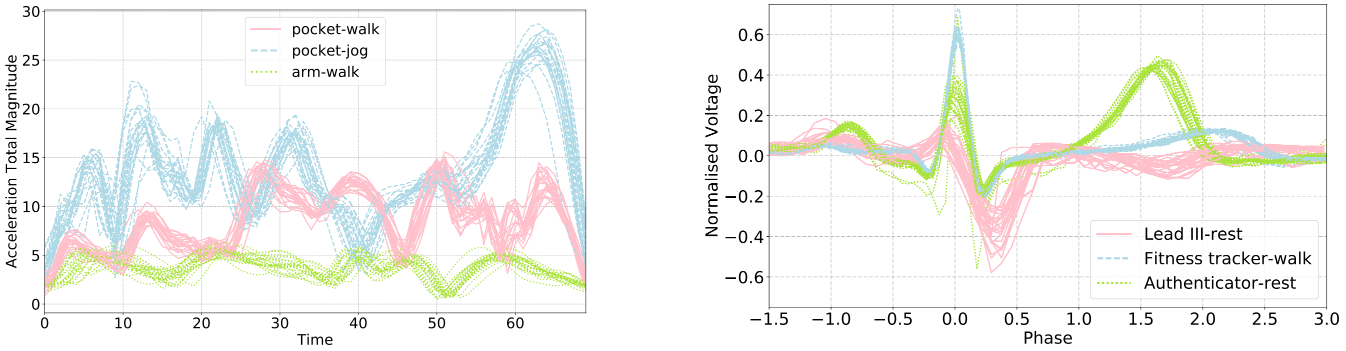
In recent years, behavioral biometrics using commodity sensors have become a popular research subject. The oldest examples are keystroke dynamics (distinctive typing patterns) and mouse movement biometrics. With the increasing prevalence of smartphones and tablets, the distinctiveness of touchscreen usage patterns has been investigated. Human gait has also been demonstrated to be distinctive, its particular appeal lies in the fact that it is easily captured through cheap accelerometers which are nowadays provided in most smartphones and smartwatches.

While these biometrics are often initially evaluated under a zero-effort threat model, the research community has recently been more focused on active attacks. Typical attacks are two-fold: (i) the attacker obtains the victim's biometric information and (ii) presents it to the authentication system. The second step can be achieved through the attacker using the system as intended while modifying their own behavior (manual imitation attack) or by using some technical contraption (robotic imitation attack). Most documented attacks have been focused on the second part of the attack (i.e., presenting previously obtained data to the authentication system). In this paper, we more thoroughly investigate challenges involved in the attacker obtaining the victim's biometric information in the first place.

Lots of attention has been given to the protection of biometric templates, not only for security but also for privacy reasons (templates are sensitive user data). As a result, most biometric authentication systems implement strong measures to protect the user's template. A classic example is using secure enclaves to store the templates (e.g., Apple Touch ID[1]). As secure elements offer good security guarantees (i.e., they are hard to bypass), attackers will likely need to obtain the victim's biometric data elsewhere.

The key challenge lies in the fact that the distribution of biometric features strongly depends not just on the user, but also on the *context* of the measurement. We use the term *context* to refer to all the factors that influence the outcome of a biometric signal measurement. Some of these factors

[1] https://support.apple.com/en-gb/HT204587



(a) Gait: total acceleration magnitude for a set of individual gait cycles produced by different contexts (for the same user).

(b) ECG: voltage of individual ECG waves produced by different contexts (for the same user).

Figure 1: Biometric signals differences across contexts.

have been already identified by Jain et al. [1]: the authors find that sensor limitations, type of task and environmental changes all influence the measurement, making it noisier. Figure 1a shows how the context of the gait biometric is expressed through the location of the sensor on the user's body (pocket or arm) and walking style (walk or jog). Similarly, Figure 1b shows the difference in the voltage of ECG waves when the measurement occurs in different contexts: depending on the placement of electrodes on the user's body (either wrist, arms and legs, or chest), or the type of activity (rest or walk). Due to these feature differences, an attacker using an arbitrary source of biometric information for their attack (akin to a replay attack in network security) will be unlikely to succeed. However, differences between contexts may be partly systematic, i.e., consistent and predictable for a large number of users.

Key Research Questions.

- How can attackers obtain biometric information without compromising the template on the authentication device?
- Are context-specific changes in features systematic and can they be predicted?
- How does this predictability impact the security of biometric authentication?

To answer these questions, we generalize the methodology of [2] to formalize an approach to automatically derive a cross-context feature mapping based on population data. The mapping then enables us to score both individual features and biometrics with regard to their predictability across contexts. Unpredictable features contribute to the overall biometric's security guarantees, as the attacker struggles to collect useful biometric information outside the context where the authentication occurs. The unpredictability score gives more information than the attack success rate, which depends on the specific implementation of a biometric system (e.g., matching algorithm, decision thresholds). Instead, this measure of security enables not just the comparison of different biometrics, but also to harden feature-sets to be more resilient against this attack. We use our methodology to assess and compare the security of five biometrics: gait, touchscreen dynamics, ECG, eye movements and mouse movements. We collect data through a user study involving 30 participants providing data for all biometrics in different contexts across two sessions. We choose contexts to reflect a variety of real-world threat scenarios.

Contributions.

- We identify a number of scenarios that enable attackers to gain access to sources of biometric information.
- We provide an analytical framework that measures to what degree biometric features can be predicted across different contexts.
- We conduct a two-sessions study on 30 users in order to collect five behavioral biometrics (gait, touch dynamics, ECG, eye and mouse movements) across a variety of contexts. We make this dataset publicly available[2].
- Based on this dataset, we use our framework to quantify the predictability of the five biometrics' features and discuss the resulting security implications.

Organization. The remainder of the paper is organized as follows: In Section 2 we summarize related work dealing with biometric authentication systems and attacks on these. Section 3 outlines the threat model we use throughout the paper. Section 4 details how our data collection methodology is built on the threat model. Section 5 describes the mathematical foundation of the cross-context mapping; the results of applying this mapping to our dataset can be found in Section 6. We conclude the paper in Section 7.

2. Related Work

Behavioral biometrics have become an increasingly popular research area, particularly sparked by the availability of a wide variety of cheap sensors in consumer electronics. In this section, we give an overview of research regarding biometric system design and attacks on these systems.

[2] https://ora.ox.ac.uk/objects/uuid:0175c157-2c9b-47d0-aa77-febaf07fca71

2.1. Biometric Authentication

Frank et al. were amongst the first to investigate the feasibility of using touchscreen input patterns for user authentication on smartphones [3]. They collect data through an image comparison game which requires users to swipe between images, with each swipe (or stroke) generating a single feature vector. The feature-set is composed of the pressure distribution across the swipe, the start and stop coordinates, the swipe's curvature as well as speed and acceleration. The authors consider (horizontal) swipes and (vertical) scrolls but argue that individual taps do not provide meaningful biometric information. Bo et al. show that a device's micro-movements generated by the user's operation of the touchscreen [4] contribute further identifying information. A recent survey of research on touchscreen input biometrics can be found in [5].

Electrocardiography (ECG) is the process of measuring the electrical activity of the heart. While the ECG waveform follows a typical pattern for all healthy humans, there are subtle differences between individuals. There is a growing body of work investigating ECG as a biometric; recent surveys can be found in [6], [7], [8]. The ECG signal is subdivided into P, Q, R, S and T waves. The main features focus on the (relative) amplitudes of the waves, their duration and the spacing between their peaks. Research on ECG biometrics has also resulted in a commercial product, the Nymi Band[3], which serves as a multi-factor authenticator.

Human gait (walking style) has garnered increased interest by the research community in recent years. Information about a person's gait can either be obtained through videos or accelerometers carried by the person. The latter is particularly compelling for continuous authentication on personal electronics, as accelerometers are near-ubiquitous in smartwatches, smartphones and a plethora of wearables. For accelerometer-based gait biometrics, the signal is typically first divided into cycles to isolate individual steps. Following that, features are obtained by dividing the reference cycle into segments, with each feature representing the acceleration within the corresponding segment [9]. Often, dynamic time warping is used to adjust for noise in the movement before template matching [9]. An overview of state-of-the-art approaches to gait recognition can be found in [10].

Optical eye trackers are available as stand-alone devices, but are also increasingly integrated in mobile devices. Tracking is typically achieved by shining a pattern of infrared light on the user's eyes and capturing the reflection of the cornea. Therefore, an eye tracker only requires a standard webcam with an attached source of infrared light, although a higher number of frames per second is needed for higher sampling rates. While eye tracking has been used in the past for medical diagnosis (for disorders such as Alzheimer's [11] and schizophrenia [12]), it has recently attracted significant interest as a biometric. Earlier work authenticates users while they are being shown controlled stimuli, such as images [13] or moving shapes [14]. Eberz et al. authenticate users while they perform standard computer tasks (reading, typing, browsing and watching videos) [15]. Their feature-set consists of temporal features reflecting short-term speed and acceleration, spatial features that measure the steadiness of the gaze and the changes of the pupil diameter. Their results show that training and operating the system on different tasks is possible, but that features show varying degrees of task dependence which leads to higher error rates.

Using distinctive mouse movements for authentication has garnered significant attention due to the near-ubiquitousness of mice in desktop environments. A survey of mouse movement biometrics can be found in [16]. Key features reflect the speed of mouse movements, their curvature and properties of mouse clicks (e.g., click duration). The click duration in particular has been shown to depend on the physical device [17].

Ref	Biometric	Type of Attack	Knowledge
[18]	Keystroke dynamics	Assisted manual imitation	Perfect
[19]	Touch dynamics	Assisted manual imitation	Perfect
[20]	Touch dynamics	Automatic (robot)	None
[21]	Touch dynamics	Automatic (robot)	Perfect
[22]	Gait	Assisted manual imitation	Perfect
[2]	ECG	Signal generator	Cross-device

TABLE 1: Overview of attacks on biometric systems.

2.2. Imitation Attacks

Typically, biometrics are evaluated under a zero-effort threat model, any successful attacks are then a result of the attacker's biometric features being sufficiently similar to the victim's template. With the growing interest in behavioral biometrics, researchers have turned towards more sophisticated attacks. These imitation (also known as *mimicry* [23]) attacks can be divided into two categories: manual imitation attacks and robotic imitation attacks. The former involves an attacker using the system as intended while modifying their own behavior to match that of the victim. For the latter, the attack is not carried out by a human but is instead automated. This approach may make it necessary to defeat any liveness detection a system may have and is also usually trivial to spot by a human observer. A summary of attacks on biometric systems is given in Table 1.

Tey et al. demonstrate a manual imitation attack against password authentication that has been hardened through keystroke dynamics [18]. As such, the user has to type in the correct password with the correct inter-key timings. The authors assume the attacker has full knowledge of the model (i.e., possesses both the password and the timing template). Users are trained to act as attackers and are provided with a training interface that gives positive and negative feedback depending on their closeness to the actual victim's timings.

Khan et al. develop a similar system to attack touchscreen input authentication [19]. They investigate two scenarios, the first one involves the attacker observing the victim's template via shouldersurfing, the second assumes perfect information gained by the attacker tricking the victim

[3] https://nymi.com

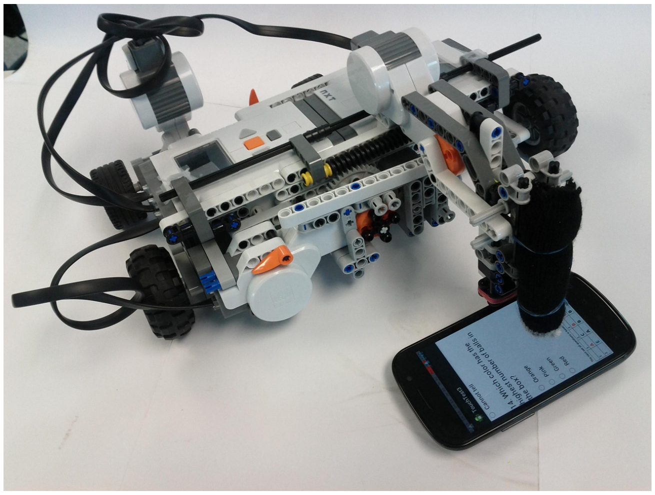


Figure 2: A robot imitating a human's touch dynamics (taken from [20]).

into using a compromised device. Similar to Tey et al.'s work, attackers are trained through an interface giving them feedback before they carry out the actual attack.

Serwadda et al. showcase a robotic imitation attack against the touchscreen biometric [20] (see Figure 2). The swipes on the victim's phone are not carried out by a human, but by a purpose-built Lego Mindstorm robot. For the original attack, the feature vectors imitated by the robot are derived from population data, rather than the specific victim. This approach significantly increased the system's false accept rate, although the baseline equal error rate is already much higher than that of related work. The authors also consider a targeted attack, for which they assume the attacker has obtained a perfect copy of (some) of the victim's feature vectors [21].

Rajesh et al. develop a manual imitation attack against the gait biometric [22]. They assume that the attacker has obtained the victim's biometric template. Using a treadmill, the attacker can modify step length, step width, speed and thigh lift while carrying out the attack. As most gait features are highly dependent on these four gait characteristics, the use of the treadmill makes the attack very effective.

The previous attacks assume that the attacker has obtained a perfect copy of the victim's template (see Table 1). However, assuming the actual template is stored securely, this may not always be a safe assumption as the attacker may only be able to sample the victim's biometric in a different context (e.g., different device or environment). Eberz et al. demonstrate an attack on ECG biometrics that considers different sources of information for the attacker, including e-health and medical devices [2]. The victim's data is injected into the authentication system by using a standard audio player as a signal generator. Their work shows that the distribution of ECG features depends on the measurement device and that the success rate of the attack drops when data is not obtained on the actual authentication device. To mitigate this, the authors propose a mapping function based on population data that accounts for device-specific feature differences. This methodology is the foundation of the cross-context mapping presented in Section 5.

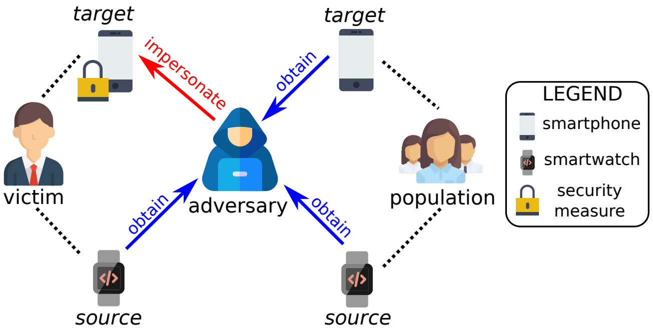


Figure 3: Example of threat scenario.

3. Threat Model

In this paper we focus on adversaries that attempt to bypass a biometric authentication system using incomplete biometric information about the victim from another *context* and combining it with population data.

Overview. Figure 3 shows an example of such a scenario. The victim is enrolled into a gait authentication system through their phone. We refer to the system and the context used by the system as *target*. The system maintains a confidence in the user's identity based on their gait patterns and allows certain sensitive operations (e.g., authorising payments) only when the confidence is above a threshold. The victim's biometric template is stored on the phone in a trusted module that cannot be accessed by the adversary.

The adversary knows that the victim uses a smartwatch that monitors their gait, for example for health or sport reasons. We refer to the context of the smartwatch as *source*. Either the smartwatch, its connected smartphone application or the wireless link are insecure and the adversary exploits the smartwatch to obtain the victim's gait data. However, as previously illustrated in Figure 1a, the smartwatch data cannot be used directly to impersonate the user at the *target* system, because of feature differences caused by the different context. Therefore, the adversary collects biometric data from a population (which excludes the victim), reproducing the *source* and *target* contexts. Using only population data, the adversary attempts to learn how to transform gait data from *source* to *target* and uses this information to transform the stolen victim's gait. The transformed data allow the adversary to impersonate the victim at *target*.

Assumptions. The victim is enrolled into a biometric authentication system (*target*). The biometric data used by the *target* system is measured in a pre-defined context (*target* context). The attacker wants to impersonate the victim at *target* system. We assume the following:

- Obtaining the victim's biometric data usable in *target* context is hard, because the devices that process *target* system data are highly protected;

- The victim uses another system that makes use of the same type of biometric data as *target*, we refer to this as *source* system. Data from *source* are more easily obtainable, but are measured in a different context (*source* context);
- The adversary can obtain biometric data from a population for *source* and *target* contexts (i.e., same sensor, task, environment, etc.);
- The adversary knows the biometric features used for recognition by *target* system, but does not know any other detail used by the recognition algorithm;
- The adversary can reconstruct biometric signals from biometric templates and can inject forged biometric data into *target* system.

It should be noted that the biometric data for *source* and *target* can either be raw biometric signals or vectors of biometric features. In fact, since the adversary knows the feature extraction algorithm used by the system, they can easily compute features from raw signals.

The adversaries may obtain population data in different ways. As an example, they could ask their friends to provide their biometric samples, or invite members of the general public for a lab study. For some biometrics, it might also be possible to use publicly available data (e.g., medical databases for ECG). Although the adversary may need to invest time and effort in collecting the population data, it is a worthwhile investment. In fact, once the universal transformation is learned, the adversary can use it to impersonate potentially anyone. In the case of local authentication the adversary will have to obtain physical access to the device and, depending on the method of injection, bypass liveness detection. On the other hand, if authentication is performed remotely (i.e., on a server), the adversary can perform the attack in a more scalable way.

In the following subsections, we motivate the threat model by presenting different scenarios for each of the biometrics. In each scenario, we outline how different factors contribute to the feature differences between contexts.

3.1. Gait

Little attention is being given to the confidentiality of accelerometer-based gait data. At the time of writing, accessing the accelerometer does not require a permission in the Android Manifest file (Android v8.0), and can be accessed directly by websites, through the DeviceMotion API. This means that adversaries might obtain control of an application (or make an application or website themselves) and silently collect data from oblivious users. Furthermore, most fitness trackers have been proven vulnerable to exploits, both in the wireless channel [24] or in the firmware [25].

With the adversary being more likely to obtain data from a fitness tracker (or a fitness application running on the smartphone), two main factors should be considered. The first one is the on-body location of the accelerometer sensor and the second one is the type of movement: either walking or running. The rationale behind the location is that different parts of the body are subject to different accelerations (e.g., arms, chest, or wrists). On the other hand, the use of fitness trackers is more popular while running than walking (e.g., to monitor work-out statistics). Attackers need to consider that running data looks extremely different from walking due to the stronger forces generated by the run and the shorter timing between steps.

3.2. Touch Dynamics

In the case of touchscreen data, there are two main ways in which the adversary could obtain users' biometric data: a malicious mobile application, or a malicious website. Adversaries could create applications that silently monitor the touchscreen inputs and trick victims (e.g., through social engineering) into installing and using these applications on their smartphones. Similarly, touchscreen data collection could be carried out on a website through simple Javascript[4].

For touch devices, we focus on the scenario where victims use at least two phones, where one of them is highly protected. An example for this scenario is where victims have a company-issued smartphone that contains company-sensitive information and is secured with different means (e.g., trusted modules, touchscreen lock, no installation of arbitrary apps allowed). In particular, the device continuously authenticates the user using touchscreen biometric while they are using sensitive applications.

The adversary will try to obtain the user biometric from a less protected device (e.g., their personal smartphone). In this case, the first factor to account for in the transformation is the dimension of the touchscreen, as these changes the span/shape of the swipe gesture. Additionally, the sampling rate of the touchscreen has a significant effect, as less fine-grained information changes the meaning of features based on a subset of the swiping gesture (e.g., initial acceleration of swipe). Other sensor data, such as pressure or area covered, might also be different in terms of scale, resolution, precision and granularity.

3.3. ECG

Similarly to gait (Section 3.1), insecurities in the communication channel or the device firmware can both be a point of attack for the adversary that is attempting to obtain ECG data. In addition, computerized medical records are often handled poorly in terms of their confidentiality. Reports show that large amounts of sensitive healthcare data are vulnerable to leakage or theft, or have already been compromised because of security lapses at hospitals, insurance companies or government agencies [26]. Adversaries may also easily obtain raw ECG signal from photos of ECG printouts, as has been shown in previous work [2].

It has been previously shown that the type and location of ECG sensors affect the ECG measurement [2] and therefore cause differences in feature distributions. Comparably to gait, with fitness trackers being more likely to be exploited, the adversary should also account for the different

[4] https://developer.mozilla.org/en-US/docs/Web/API

Biometric	Factors	Considered Scenarios	Devices
Gait	Activity Sensor Location Input Device	walking, running arm, chest, hand, pocket, wrist smartphone, smartwatch, fitness tracker	BLU VIVO 6, Movisens ekgMove, Garmin Vivoactive HR
Touch dynamics	Input Device	low-, mid-, high-end phone	TTSim M5 Smart, Motorola MotoG3, BLU VIVO 6
ECG	Sensor Type Activity	mobile monitor, medical monitor, fitness tracker, authenticator resting, walking, running	AliveCor KardiaMobile, Heal Force Prince 180B, Movisens ekgMove, Nymi Band
Eye movements	Task Calibration	reading, watching video, writing, browsing calibrated, uncalibrated	SMI RED500
Mouse movements	Input Device	trackpad, mouse	MSI GT72 6QE Dominator Pro G trackpad, Dell Laser USB mouse

TABLE 2: Factors of feature distribution differences considered for each biometrics and devices used for the measurements.

ECG behavior due to the activity performed by the user during the measurement. The ECG signal significantly changes when the user is exercising, both due to the physical exertion and noise introduced by imperfect electrode connection.

3.4. Eye Movements

The popularity of eye-tracking is increasing and a number of consumer electronics are equipped with eye trackers. With more services implementing eye-tracking, adversaries can use these services to obtain eye movement data (e.g., hijacking browsers and using a Web API, or exploiting application weaknesses). Additionally, we consider the threat of the user being tricked into using an attacker-controlled machine which is equipped with a covert eye tracker.

For eye movements, it has been shown that gaze data strongly depend on the type of task performed by the user (e.g., reading, writing, browsing [27]). Since the adversary can not easily force the user into performing a specific task, they might need to adapt the victim's data to the task that is used for authentication. Additionally, eye trackers need to be calibrated before use to provide accurate data. Since it would be considerably more difficult to trick the user into calibration (as this procedure would raise suspicion), we assume that the attacker only possesses data from a device that is not calibrated for the victim.

3.5. Mouse Movements

As mentioned in Section 3.2, collection of mouse movements data can easily take place on the Web, where it has been shown that mouse tracking is common-place [28]. In order to obtain the victim's data, adversaries may create websites, or hijack existing ones. It could also be possible for the adversary to highjack the victims' browsers (e.g., by installing malicious extensions [29]).

As users interact (and browse) with an increasing number of devices, the adversary needs to account for the different interactions that happen depending on the device hardware. Previous work shows that changing the pointing device hardware causes fluctuations in the measured users behavior, enough to significantly degrade the recognition performance [17]. Using these observations, we decide to consider the extreme case where the pointing device is either a mouse or a trackpad. This fits well the scenario where mouse data collection happens remotely, that is the most likely to occur online (as mentioned above).

4. Experimental Design

In order to evaluate the threat model motivated in the previous section, we conduct a study where we collect participants' biometrics for each of the five biometric modalities. The study is designed to reflect the scenarios presented in the threat model (Section 3). For all biometrics measurements, we stick to state-of-the-art common practices. In the following, we describe the details of the study and briefly comment the processing methodologies that we adopt.

4.1. Study Outline

The study consists of two separate but identical sessions which are at least 5 and not more than 30 days apart. In each session, participants undergo a series of tasks designed to collect their biometric traits for a specific context. A single session lasts approximately one hour and 45 minutes. In Table 2 we report all the feature difference factors that we accounted for in the analysis and the devices used for the measurements. In the remainder of this section, we present the details of the study procedure for each biometric.

Mouse Movements. The first task is carried out on a laptop to collect mouse data [30]. Participants are shown a grid of rectangles and click on the rectangle that contains a picture. After the user clicks, the picture moves to another rectangle and users click on this new rectangle. The task ends after 250 total clicks and is repeated with the trackpad.

Eye Movements. The participant is then requested to complete five different tasks on a laptop equipped with an eyetracker. The study is carried out in a lab in controlled lighting conditions (blinds closed and light switched on). We take our tasks from the experimental design of [15]: reading, writing, watching a movie trailer, browsing and watching an educational video. Each task continues for 3 minutes before the next one automatically starts. Differently from [15], we include two different videos to account for the number of scene changes that directly influence the participant's gaze: the movie trailer contains lots of fast-paced scene changes, while the educational video does not. At the end of the session, the five tasks are repeated on an uncalibrated eyetracker. To account for the users getting used to the tasks when they repeat them, we randomly swap the order of the calibrated and uncalibrated tasks.

Touch Dynamics. Afterwards, the participant uses a smartphone to complete a "spot the difference" task (similarly to [3]). The smartphone shows two images which contain subtle differences between each other and the user attempts to find them. Only one image is shown on the smartphone at a time and the user swipes (either to the left or to the right) to see the other image. The task lasts 3 minutes in total and is repeated three times, each time with a different phone and a different pair of images. To avoid bias generated by the selection of images and users acclimating to the task, for each user we randomize the order of the phones.

ECG. Then, the participants ECG is monitored for a set of devices: an authenticator, a mobile ECG monitor attached to a smartphone and a medical ECG monitor. For the ECG monitor measurement, we collect the palm measurement using the built-in electrodes and use an external 3-lead ECG cable with disposable electrodes, to obtain Lead I, Lead II and Lead III [31]. Additionally, at the beginning of the session, participants wear a chest-strap fitness tracker that monitors their ECG and gait data throughout the session.

Gait. Finally, the participant goes for a short walk and a subsequent run in a nearby park (around 700 meters each). During this time, five different sensors monitor the participant's gait pattern: three smartphones (placed on left arm, right front pocket and held in the left hand), a smartwatch worn on the left wrist and the fitness tracker mentioned above. The fitness tracker also monitors ECG during the walk and the run.

Participant Recruitment. We recruited a total of 30 (11 female, 19 male) participants through local announcements and social media. Participants were compensated for their time and inconvenience. This study was reviewed by and obtained clearance from the Inter-Divisional Research Ethics Committee of the University of Oxford, reference number R50977/RE001.

4.2. Feature Extraction

We adopt state-of-the-art common practices for biometric data processing and feature extraction. Table 3 reports the papers we used. For ECG and gait we chose to use preprocessing steps to allow us to isolate the individual signals (single heartbeat and single gait cycle, respectively), rather than frequency domain analysis. The rationale behind this choice is that feature representation based on frequency domain does not have a direct and understandable meaning, while providing similar (if not weaker) performance results.

Biometric	Paper(s)	Description
Gait	M. Derawi et al. [9]	magnitude of acceleration features (based on cycle detection)
Touch dynamics	M. Frank et al. [3]	pressure, spatial, speed and acceleration features
ECG	A. Fratini et al. [7]	temporal, amplitude, morphology features (based on fiducial points)
Eye movements	S. Eberz et al. [15]	pupil, temporal and spatial features
Mouse movements	N. Zheng et al. [32] A. Weiss et al. [30]	stroke curvature, speed and acceleration features

TABLE 3: Description and original paper of the preprocessing and feature extraction methodologies used for each biometric.

For gait, we ignore the use of dynamic time warping, as this is only necessary during template matching and does not have an effect on the raw signal behavior. Due to limited space, we report all the individual features and their importance based on Relative Mutual Information (RMI) in Appendix A. The details of the feature extraction for each biometric can be found in the cited papers.

5. Mapping Methodology

In this section, we describe the methodology used to derive the cross-context mapping and discuss how the mapping is combined with feature distinctiveness. Finally, we discuss the evaluation methodology and how these results should be interpreted.

5.1. Cross-Context Mapping

We generalize the cross-device mapping approach introduced in [2] to cross-context mappings and transformations chosen from a parametrized family of functions. Given a source and target context, for each user we have a set of feature values computed in the source context and a set of feature values computed in the target context. Notice that source and target features are computed *independently* from each other (i.e., in different experiments). This is different from function regression in which inputs and output values are assumed to be measured in a paired way; rather, we have a set of input values and a set of output values measured in independent experiments.

For each pair of the source-target context, we compute a mapping on each biometric feature. Intuitively, this cross-context mapping works by optimizing the intra-user statistical similarity between the feature values sampled from the source context and those sampled from the target context. More formally, given a context and a feature, we associate to each user a random variable. The latter models the experiment of observing specific feature values, for each user. The cross-context mapping then transforms each source context random variable to maximize statistical similarity to its corresponding target context random variable. The final output of the estimation, for a pair of the source-target context, is a set of functions (one per feature) that maps the values of features measured in source to the target.

Problem Setting. Let $\{u_i\}_{i=1,\dots,n}$ be the set of users from the population for whom we have observations for both the source s and the target t contexts, which we refer to as $\{x_{u_i,j}^{(s)}\}_{j=1,\dots,n_{u_i}^{(s)}}$ and $\{x_{u_i,j}^{(t)}\}_{j=1,\dots,n_{u_i}^{(t)}}$; and let v be the victim for whom we have observations only from the source context s, that is, $\{x_{v,j}^{(s)}\}_{j=1,\dots,n_v^{(s)}}$. For each user u_i in the population and for the victim v, let $X_{u_i}^{(s)}$, $X_v^{(s)}$ and $X_{u_i}^{(t)}$, $X_v^{(t)}$ be the random variable associated to a specific feature from the source and the target contexts. respectively. We seek an optimal transformation function f^* of the source random variables such that, for $i = 1, \dots, n$, $f^*(X_{u_i}^{(s)})$ and $X_{u_i}^{(t)}$ are, statistically speaking, similar. We then use $f^*(X_v^{(s)})$ as an estimation of the unknown target random variable for the victim, i.e., $X_v^{(t)}$. In other words, f^* transforms each value of the source feature to be as close as possible to what would be observed for the target feature.

Cross-Context Mapping Estimation. As in function regression in finite dimensional vector spaces, the estimation of f^* is composed of three phases: (i) the definition of a parameterized family of functions $\{f_\theta\}_{\theta\in\Theta}$ which to optimize for, (ii) the definition of an error function for each f_θ and (iii) the solution of an optimization problem in which the overall error is minimized with respect to the generic transformation function f_θ.

For a generic user u in the population, we evaluate the dissimilarity between the target random variable $X_u^{(t)}$ and the transformed source random variable $f_\theta(X_u^{(s)})$ as the statistical distance between two cumulative density functions associated with the two random variables[5]. The rationale is that, if the two variables have the same distribution, then they are indistinguishable by the template matching algorithm. Namely, let $F_{f_\theta(X_u^{(s)})}$ and $F_{X_u^{(t)}}$ be the two cumulative density functions, we define the error ϵ that the function f makes for user u as:

$$\epsilon_{f_\theta}(u) = d\left(F_{f_\theta(X_u^{(s)})}, F_{X_u^{(t)}}\right), \tag{1}$$

where d is a generic statistical distance between cumulative density functions (discussed in the following paragraph). The optimal function can hence be defined as the transformation function f_{θ^*}, where θ^* is the vector of parameter values that minimizes the across-users average error, that is:

$$\theta^* = \arg\min_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} d\left(F_{f_\theta(X_{u_i}^{(s)})}, F_{X_{u_i}^{(t)}}\right). \tag{2}$$

In the following, we reformulate the estimation problem of Equation 2 using a specific distance function and Θ.

Optimization Problem. We define the distance d to be the L^2 distance between functions, that is:

$$d\Big(F_{f_\theta(X_u^{(s)})}, F_{X_u^{(t)}}\Big) = \sqrt{w\int_{\mathbb{R}}\Big(F_{f_\theta(X_u^{(s)})}(\xi) - F_{X_u^{(t)}}(\xi)\Big)^2 d\xi}. \tag{3}$$

Previous work shows that the precise choice of distance measure has little influence in cross-device settings [2]. Factor w in Equation 3 is a factor used to normalize the distance d in the interval $[0, 1]$ (details on the computation of w are given in Appendix B). Further, we tweak the objective function of Equation 2 to be robust against noisy estimations for the distributions of particular users. Namely, let $I = \{1, \dots, n\}$, then for each function f_θ we define a subset of the user population indexes $I_{f_\theta} \subseteq I$ as follows: (i) we compute the distances $d(F_{f_\theta(X_{u_i}^{(s)})}, F_{X_{u_i}^{(t)}})$ for all users, (ii) we iteratively apply the Grubbs test to detect a subset of outlier indexes $I_{f_\theta}^o$ among these distances and (iii) we remove the users' labelled as outliers, $I_{f_\theta} = I \setminus I_{f_\theta}^o$. In doing this, we set the test significance level to 0.1 and use 10% as the maximum percentage of outliers included in $I_{f_\theta}^o$.

Implementing outliers detection in the error function of Equation 2, we obtain a non-linear optimization problem in the real-valued vector of parameters θ, which we solve by using a pattern search optimization algorithm [33]:

$$\begin{aligned} &\underset{\theta}{\text{minimize}}\ \frac{1}{|I_{f_\theta}|}\sum_{i\in I_{f_\theta}}\sqrt{w\int_{\mathbb{R}}\Big(F_{f_\theta(X_{u_i}^{(s)})}(\xi) - F_{X_{u_i}^{(t)}}(\xi)\Big)^2 d\xi} \\ &\text{subject to } \theta \in \Theta, \end{aligned} \tag{4}$$

where $|I_{f_\theta}|$ is the cardinality of the index set I_{f_θ}. We refer to Appendix B for the definition of the feasible parameter region Θ. Naturally, it is not practical to test the performance of every conceivable (mapping) function. Due to their simplicity, linear functions offer good computational performance which is particularly important for larger population sizes. In the analysis of Section 6 we therefore consider linear functions f_θ[6]. We have also performed the analysis using polynomials of degree two, three and four, but none of these provided results significantly different from the linear function f_θ (all $p > 0.01$ for one-tailed Wilcoxon rank sum tests, average error decrease: $\sim$0.002, average relative error decrease: $\sim$2.6%). Full results are discussed in Appendix C.

Unpredictability Score. In order to evaluate the effectiveness of the mapping, we measure the prediction error on a per-feature base. Let v be a victim user, $\{u_i\}_{i=1,\dots,n}$ a population of users and g_j the j-th feature used by the biometric algorithm. For feature g_j, we compute the optimal

[5] Using the set of target observations $\{x_{u,j}^{(t)}\}_{j=1,\dots,n_u^{(t)}}$ and transformed source observations $\{f_\theta(x_{u,j}^{(s)})\}_{j=1,\dots,n_u^{(s)}}$ we compute the empirical cumulative density function for each random variable using the Kaplan-Meier estimate.

[6] Note that linear f_θ still poses a general non-linear optimization problem as the function $\theta \mapsto F_{f_\theta(X_{u_i}^{(s)})}(\xi)$ is still non-linear (and non-convex).

cross-context mapping $f_{\theta^*}^{(g_j)}$ (using the population) and the prediction error for the victim source observations to the victim target observations as $\epsilon_{f_{\theta*}}^{(g_j)}(v)$. This gives an *unpredictability* score U for feature g_j and victim v in the source-target context transformation:

$$U_v^{(g_j)} = \epsilon_{f_{\theta^*}^{(g_j)}}(v)\,. \tag{5}$$

Following on from Equation 3, we know that the error $\epsilon_{f_{\theta^*}^{(g_j)}}(v)$ lies in the interval $[0, 1]$. A small value of $U_v^{(g_j)}$ implies that for feature g_j the cumulative functions of the victim's transformed source random variable and of the target random variable are almost overlapping. This means that (for the *j*-th feature) the cross-context mapping approach is able to accurately map observations from the source context to samples from the target context (the differences are systematic). On the other hand, a value of $U_v^{(g_j)}$ close to 1 implies that for feature g_j the transformed feature values from source random variable and from target random variable have highly non-overlapping distributions. This means that the differences between the *j*-th feature values in the source and target contexts cannot be systematically predicted in this way.

5.2. Weighted Score

Following on from the previous section, we know that we obtain an unpredictability score for each feature in the feature-set. We want to aggregate this score to the level of the whole biometric modality (across the features), so that it provides an idea of the resilience of a particular biometric to this transformation. A simple average of the unpredictability score for each feature is not reasonable, as features contribute differently to the recognition. For example, if a non-distinctive feature is very predictable, it might have a significant negative influence on the overall score. This is not the desired effect, as an attacker would gain very little by correctly predicting that feature.

RMI Weights. We weight features based on Relative Mutual Information (RMI). To avoid problems with the choice of the number of bins (that may introduce bias in the mutual information), we adopt the non-parametric RMI computation of Ross [34]. In this approach, mutual information is computed based on the relationship between a data point's neighbours and its class neighbours. We weight each feature mapping result with the feature's RMI and obtain an aggregated score that accounts for feature distinctiveness this way.

Formally, given the set of features for a biometric $\{g_j\}_{j=1,\ldots,m}$, the victim user v and each feature RMI value $\{r_j\}_{j=1,\ldots,m}$ we compute a RMI-weighted unpredictability score W_v:

$$W_v = \frac{\sum_{j=1}^{m}(\epsilon_{f_{\theta^*}^{(g_j)}}(v) \cdot r_j)}{\sum_{j=1}^{m} r_j}. \tag{6}$$

Score Interpretation. The weighted unpredictability score W_v of a biometric modality (Equation 6) depends on the scores of the individual features, with distinctive features contributing more to it. It should be noted that the score itself does not directly correspond to a certain success rate of an actual attack, because the cross-context mapping effectiveness also depends on the specific template matching algorithm and false accept and false reject rates thresholds. The main advantage of the unpredictability score lies in its comparative capability, rather than in being an absolute scale. The score can be used to ***compare different biometrics***, with biometrics with higher unpredictability scores across all sources being judged more secure. Similarly, a system developer can use the scores to ***identify vulnerable target contexts***. For example, a biometric might exhibit low unpredictability scores on specific devices (e.g., due to lower quality sensors). In that case, a developer could change the classifier's decision threshold to account for the increased danger of cross-context attacks.

Lastly, individual feature unpredictability scores $U_v^{(g_j)}$ can be a driving factor in the ***selection and engineering of features***. Higher security can be achieved both by changing the definition of features and by modifying sensor hardware (e.g., by making it less similar to common source contexts).

5.3. Evaluation Methodology

Cross-Validation. For the evaluation of the cross-context mapping, we operate in a leave-one-out cross-validation fashion. At each step i, we consider one user u_i as the victim and we use the remaining 29 users as the population. With the population, for each feature, we compute the optimal cross-context mapping $f_\theta*$ and the prediction error for the victim source observations to the victim target observations. We obtain $U_v^{(g_j)}$ (Equation 5) and W_v (Equation 6) this way. This step is repeated for each user. If not otherwise specified, the results shown are averages of unpredictability scores over the users in our dataset. The RMI is computed on the feature distribution of the population obtained in the first session, for the target context.

Considered Scenarios. In the evaluation, we select a set of sources for each biometric and consider the scenario where the adversary has the information from an individual source, or for the full set of sources (*all*). In the second case, the adversary uses the source with the best performing cross-context mapping (lowest unpredictability) for *each feature*. This scenario constitutes the strongest attacker since some sources may be useful to predict some features but not others. Additionally, we consider two different time scenarios: same session and cross session. The former represents the case in which the victim's source and target data are collected in the same session, which leads to greater similarity. In the latter, the victim's source data were collected in a different session than the victim's target data. Intuitively, this reflects the case of the attacker's source data being older or newer than the victim's template.

Biometric contexts	Same Session avg (min, all)	Cross Session avg (min, all)
ECG		
target: *Authenticator-rest*	**.09 (.07, .06)**	**.12 (.09, .08)**
- Lead I-rest	.075 ± .010	.093 ± .014
- Lead II-rest	.106 ± .011	.128 ± .015
- Lead III-rest	.114 ± .008	.144 ± .014
- Palm-rest	.080 ± .007	.110 ± .010
- Mobile-rest	.075 ± .007	.092 ± .005
- Fitness tracker-rest	.104 ± .010	.134 ± .012
- Fitness tracker-walk	.100 ± .012	.123 ± .017
- Fitness tracker-jog	.103 ± .011	.122 ± .017
Eye movements		
target: *Calibrated*	**.08 (.07, .07)**	**.10 (.09, .09)**
- Intra task-uncalibrated	.068 ± .014	.089 ± .023
- Cross task-uncalibrated	.084 ± .017	.103 ± .023
Mouse movements		
target: *Mouse*	**.07**	**.07**
- Trackpad	.068 ± .011	.071 ± .010
Touch dynamics		
target: *Mid-end phone*	**.08 (.07, .07)**	**.08 (.08, .07)**
- Low-end phone	.084 ± .009	.082 ± .008
- High-end phone	.071 ± .008	.075 ± .009
Gait		
target: *Pocket phone-walking*	**.15 (.15, .13)**	**.14 (.14, .13)**
- Smartwatch-walk	.155 ± .016	.144 ± .020
- Hand phone-walk	.154 ± .021	.145 ± .019
- Smartwatch-jog	.148 ± .019	.141 ± .018
- Cheststrap-jog	.154 ± .019	.144 ± .020
- Arm phone-jog	.156 ± .020	.146 ± .021

TABLE 4: Unpredictability score, for data from the same and cross session. Rows in bold report the aggregated score, introduced in Section 5.3. For each source we also show the 95% confidence intervals computed over the unpredictability scores of individual users.

6. Results

In this section we present the results of our analysis. We first explain the choice of the source and target contexts and present high-level results. Afterwards, we show a feature-level analysis and discuss the effect of the population size.

6.1. Context Choice

In order to present data in a readable way, we select a subset of target and source contexts, following the most relevant attack vectors presented in the threat model. Of the 30 possible target contexts coming from our experimental design (see Table 2), we select five possible targets (one for each biometric) and a number of representative sources for each of them. The chosen contexts are the following:

- **Gait** – *Pocket phone-walk*: we select the pocket phone with walking activity as target. We consider five different contexts: *Smartwatch-walk*, *Hand phone-walk*, *Smartwatch-run*, *Chest strap-run* and *Arm phone-run*.
- **Touch dynamics** – *Mid-end phone*: the middle-end phone represents the reasonable choice, as it allows us to measure the effect of using higher and lower quality devices as sources.
- **ECG** – *Authenticator-rest*: the *Authenticator* (Nymi band) uses ECG for authentication purposes and therefore represent an ideal target. All the remaining ECG sensors are considered as the sources, including the different measurements obtained with the medical monitor: *Lead I*, *Lead II*, *Lead III*, *Palm*.
- **Eye movements** – *Calibrated*: all the calibrated tasks are considered as target. We consider only uncalibrated data as the source and separate between uncalibrated data coming from the same task (e.g., *Uncalibrated-reading* to *Calibrated-reading*) and uncalibrated data coming from different tasks (e.g., *Uncalibrated-writing* to *Calibrated-reading*).
- **Mouse** – *Mouse*: we select *Mouse* as a target and will use *Trackpad* as source.

Hereafter, results will refer to these target contexts. Additional results are reported in Appendix A.

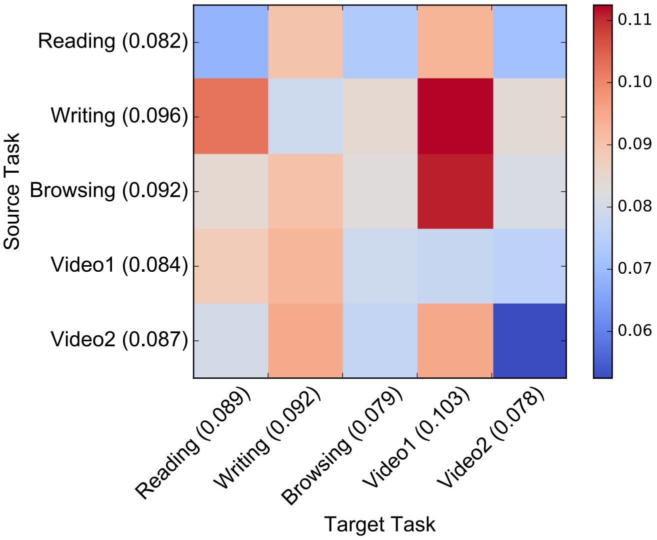
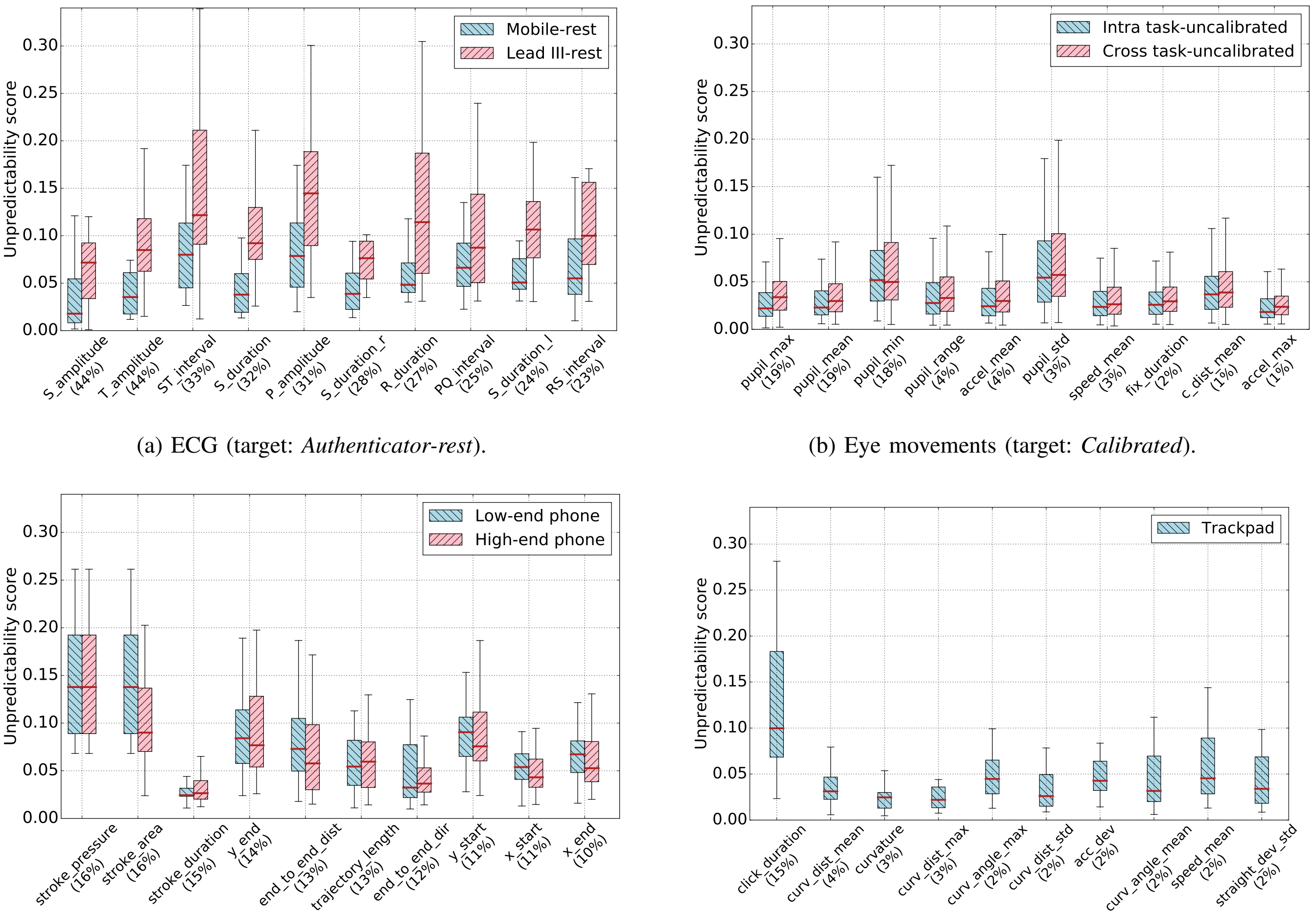


Figure 4: Unpredictability score for different combinations of source and target tasks for the eye movement biometric. Numbers in brackets are the average scores for the respective source or target. Intuitively, using data collected during the same task yields the lowest unpredictability scores (shown on the diagonal).

6.2. Biometrics Overview

In Table 4 we report the resulting RMI weighted scores for each target and source context considered in Section 6.1. The first rows report the aggregated results over the sources: *average*, *minimum* and *all* weighted score (see Section 5.3). In Table 4, we can see that biometrics rank differently in terms of unpredictability. The table shows that ECG and gait are on average more resilient to the cross-context transformation, in both the same session and cross session scenarios. Gait in particular is very resilient to cross-context attacks, with an unpredictability score two times higher compared to touch dynamics, eye and mouse movements.



(a) ECG (target: *Authenticator-rest*).

(b) Eye movements (target: *Calibrated*).

(c) Touch dynamics (target: *Mid-end phone*).

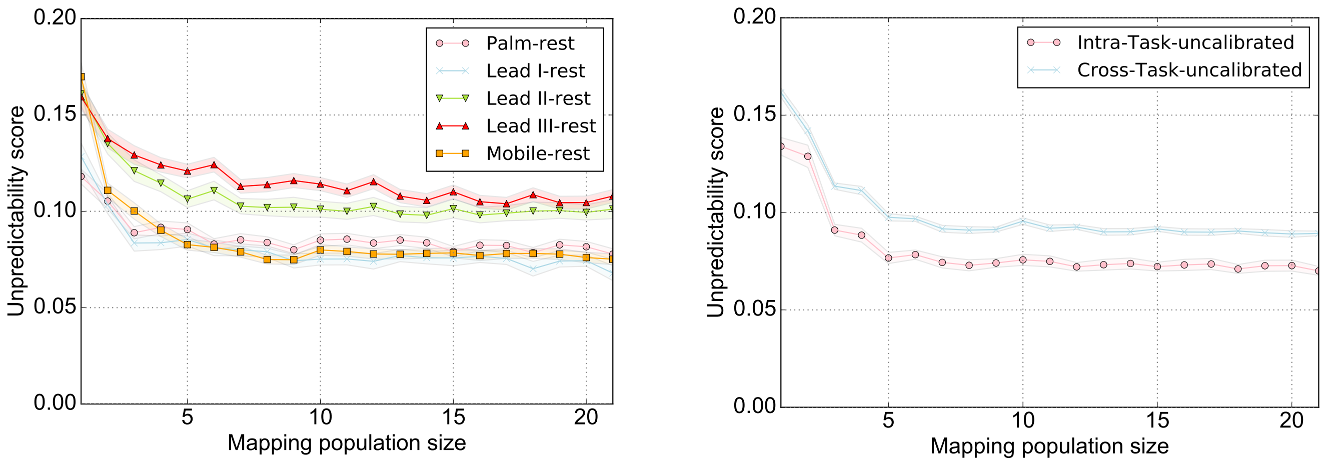
(d) Mouse movements (target: *Mouse*).

Figure 5: Unpredictability score of the top-ten RMI ranked features for the ECG (Fig. 5a), eye movements (Fig. 5b), touch dynamics (Fig. 5c) and mouse movements (Fig. 5d) biometric. Features are sorted by RMI (descending from left to right). RMI is reported in percentage on the x-axis label.

This means that the different placement of the sensors provide poor information about the gait signal as measured in other contexts. Comparatively low results are obtained for eye movements, touch dynamics and mouse. Most of these biometric features are easily and consistently mapped across source contexts (see the discussion in Subsection 6.3). For the eye movements biometric there are also differences depending on the respective source and target task, shown in Figure 4. Naturally, intra-task mappings produce the lowest unpredictability score (as the only difference is the lack of calibration for the source task), while cross-task mappings perform particularly poorly for some combinations. The results show that an attacker could gain a significant advantage if they are able to choose the source task freely.

Comparing the average, minimum and all score we can see that: (i) by selecting the appropriate source context the adversary can expect an improvement of $\sim 10\%$ on average, that is, from *average* to *minimum* score (consistent across same and cross session scenarios); (ii) by combining information from several sources the adversary might obtain a further improvement up to $\sim 15\%$ (again consistent across same and cross session scenarios), that is, from *minimum* to *all* score. This means that it might be worthwhile for an adversary to obtain biometric information over a higher number of sources and selectively choose to map individual features from whichever source provides the lowest unpredictability score for that feature.

The results show that same-session scores are lower compared to cross-session scores for ECG and eye movements in particular. As a result, an attack would appear to be more likely to succeed if very recent data (as in the same-session experiment) is used. However, the authentication system itself has to cope with the (lack of) time stability which causes this difference. Most likely, this will be achieved through either periodic retraining or continuous template updating. While template updating will make false rejects as a result of increasing time distance less likely, it will also enable the attacker to use older data for the attack.



(a) ECG.

(b) Eye movements.

Figure 6: Effect of the mapping population size on unpredictability scores.

6.3. Feature Analysis

In order to understand to what extent individual features contribute to the overall score, we analyze them separately. We report in Figure 5 boxplots for the raw (non-weighted) cross-context feature unpredictability scores. Each box shows the unpredictability score for a single feature from the source context to the target context. Features are ordered by decreasing RMI on the x-axis and the RMI value is reported on the x-axis. For conciseness, for each biometric, we only show a couple of meaningful sources and present just the top-ten RMI-ranked features, as these are the ones that contribute the most to the weighted score.

ECG. We notice that the type of sensor used as the source has a significant impact on the weighted unpredictability score (confirming the results of [2]). In Figure 5a we can see how *Mobile* consistently outperforms *Lead III* for each feature. This can be explained by closer similarity of the ECG signal when measured at the extremity of the subject's arms (true for *Lead I*, *Mobile* and the target *Authenticator*) compared to for example the *Lead III* measurements, which measures voltage potential between the left arm and left leg. The differences in predictability for different sources shown in Table 4 and Figure 5a highlight that ECG-based authentication might still be secure if the adversary steals ECG data from dissimilar contexts, but becomes less secure the easier it is to obtain data from similar contexts. Hand-based measurements are convenient and common (as shown by the popularity of e-health devices), this highlights the danger of using the same type of measurement for authentication.

Eye movements. Figure 5b shows how most eye movements features are highly predictable, both pupil-based and speed- or acceleration-based ones. The boxplot additionally shows how *Intra task* consistently provides relatively lower unpredictability than *Cross task*, which show that each task produces feature changes that depend on the user. Our threat model considers the case of the victim using a compromised machine with a covert eye tracker (see Section 3). The results show that if the attacker can choose the task on this machine freely (i.e., close to that on the authentication machine), he will obtain more useful data.

Touch dynamics. In Figure 5c we can see that *High-end* phones provide slightly lower unpredictability scores compared to *Low-end* phones. The low result of *stroke_duration* shows that such feature is easily predictable across devices. This is intuitively explained with users adjusting the length of their swipes to the size of the touchscreen. In a feature selection scenario, a system designer might reasonably decide to drop *stroke_duration* from the feature-set. In fact, even if the feature has a decent distinctiveness, it is extremely predictable compared to other similarly distinctive features. Overall, it is evident that the lower-end phone is a less useful source of biometric information. This is mainly due to less precise sensors (i.e., lower sampling rate and resolution), which particularly affects acceleration features (low touchscreen sampling rate) and area covered (low resolution). Conversely, this shows that high-quality sensors can improve the security of an authentication device.

Mouse movements. Figure 5d reports on average low unpredictability results for most mouse movements features. Curvature-based features in particular seem highly predictable, while not carrying significant distinctiveness (they might be dropped in a security-critical scenario). However, the plot shows a high mean and standard deviation for *click_duration*. This is due to the the trackpad API returning a coarse-grained click information, less sensitive than that returned by the mouse. Conversely, if source and target were switched, this feature would be very easy to predict as the set of valid target values would be small. This example highlights that more accurate sensors with higher resolution can thwart attacks coming from lower-quality data sources.

6.4. Population Size Analysis

Collecting a large number of (pairs of) biometric samples to train the cross-context mapping is a considerable effort. While it is possible to use publicly available datasets (see Section 3), this data may not always be available for the victim's target context (e.g., when the victim uses an unusual device). As such, it is important to know how large (in terms of number of users) the population has to be to produce acceptable results. Figure 6 shows the relationship between the number of users in the population and the average score of the resulting cross-context mapping. All biometrics show an initial sharp drop in the score and exhibit diminishing returns beyond a population of size of 10. These results show that most of the cross-context mapping's predictive power can be achieved with a relatively small population. In addition, Figure 6 suggests that the sample size of our study (30 participants) is large enough to demonstrate differences between individual features, contexts and biometrics.

7. Conclusion

In this paper, we have presented an analytical framework that allows us to measure the unpredictability of biometric features across different contexts. We define the notion of an *unpredictability* score, which can be calculated both for individual features and complete biometrics. The score provides fine-grained information about the resilience of biometric systems against cross-context attacks and can be used to: (i) compare biometric systems, (ii) identify vulnerable target contexts and for (iii) the selection and engineering of features. The framework is based on computing a mapping between a source and target context, where the mapping is derived from population data.

Our results demonstrate that the five biometrics evaluated in this paper show different degrees of resilience to cross-context attacks. In particular, we showed that ECG and gait are up to twice as unpredictable across contexts compared to touch dynamics, mouse and eye movements. Our analysis highlights particularly predictable features and suggests that some of can be reasonably dropped from the feature-set to achieve greater security against this attack. Furthermore, our data suggests that improving the quality of the biometric sensor improves the resilience of the authentication system. The fact that some contexts are more useful than others for the prediction shows that the sources of biometric information potentially available to an attacker need to be an integral part of any biometric threat model.

Acknowledgments

This work was supported by a grant from Mastercard, the Engineering and Physical Sciences Research Council [grant numbers EP/M50659X/1, EP/N509711/1] and the EU's Horizon 2020 under the Marie Skłodowska-Curie grant [No 722022]. The ekgMove sensors used in the experiment were kindly provided by movisens.

References

[1] A. K. Jain, K. Nandakumar, and A. Ross, "50 years of biometric research: Accomplishments, challenges, and opportunities," *Pattern Recognition Letters*, vol. 79, pp. 80–105, 2015.

[2] S. Eberz, N. Paoletti, M. Roeschlin, A. Patan, M. Kwiatkowska, and I. Martinovic, "Broken hearted: How to attack ecg biometrics," in *24th Annual Network and Distributed System Security Symposium*, 2017.

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

[4] C. Bo, L. Zhang, X.-Y. Li, Q. Huang, and Y. Wang, "Silentsense: Silent user identification via touch and movement behavioral biometrics," in *Proceedings of the 19th annual international conference on Mobile Computing & Networking*, pp. 187–190, ACM, 2013.

[5] P. S. Teh, N. Zhang, A. B. J. Teoh, and K. Chen, "A survey on touch dynamics authentication in mobile devices," *Computers & Security*, vol. 59, pp. 210–235, 2016.

[6] F. Sufi, I. Khalil, and J. Hu, "ECG-based authentication," *Handbook of information and communication security*, pp. 309–331, 2010.

[7] A. Fratini, M. Sansone, P. Bifulco, and M. Cesarelli, "Individual identification via electrocardiogram analysis," *BioMedical Engineering OnLine*, vol. 14, no. 1, p. 78, 2015.

[8] F. Agrafioti, J. Gao, and D. Hatzinakos, "Heart biometrics: Theory, methods and applications," in *Biometrics*, InTech, 2011.

[9] M. O. Derawi, C. Nickely, P. Bours, and C. Busch, "Unobtrusive user-authentication on mobile phones using biometric gait recognition," in *Proceedings of the 6th International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 306–311, IEEE, 2010.

[10] T. T. Ngo, Y. Makihara, H. Nagahara, Y. Mukaigawa, and Y. Yagi, "The largest inertial sensor-based gait database and performance evaluation of gait-based personal authentication," *Pattern Recognition*, pp. 1–10, 2013.

[11] A. Jones, R. Friedland, B. Koss, L. Stark, and B. Thompkins-Ober, "Saccadic intrusions in alzheimer-type dementia," *Journal of neurology*, vol. 229, no. 3, pp. 189–194, 1983.

[12] B. A. Clementz, J. A. Sweeney, M. Hirt, and G. Haas, "Pursuit gain and saccadic intrusions in first-degree relatives of probands with schizophrenia," *Journal of abnormal psychology*, vol. 99, no. 4, p. 327, 1990.

[13] V. Cantoni, C. Galdi, M. Nappi, M. Porta, and D. Riccio, "Gant: Gaze analysis technique for human identification," *Pattern Recognition*, vol. 48, no. 4, pp. 1027–1038, 2015.

[14] Z. Liang, F. Tan, and Z. Chi, "Video-based biometric identification using eye tracking technique," in *IEEE International Conference on Signal Processing, Communication and Computing*, pp. 728–733, IEEE, 2012.

[15] S. Eberz and K. B. Rasmussen, "Looks like eve: Exposing insider threats using eye movement biometrics," *ACM Transactions on Privacy and Security*, vol. 19, no. 1, 2016.

[16] K. Revett, H. Jahankhani, S. T. Magalhães, and H. M. Santos, "A survey of user authentication based on mouse dynamics," *Global E-Security*, pp. 210–219, 2008.

[17] Z. Jorgensen and T. Yu, "On mouse dynamics as a behavioral biometric for authentication," in *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security*, pp. 476–482, ACM, 2011.

[18] C. M. Tey, P. Gupta, and D. Gao, "I can be you: Questioning the use of keystroke dynamics as biometrics," in *20th Annual Network and Distributed System Security Symposium*, pp. 1 – 16, 2013.

[19] H. Khan, U. Hengartner, and D. Vogel, "Targeted mimicry attacks on touch input based implicit authentication schemes," in *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*, pp. 387–398, ACM, 2016.

[20] A. Serwadda and V. V. Phoha, "When kids' toys breach mobile phone security," in *Proceedings of the 2013 ACM SIGSAC conference on Computer & Communications Security*, pp. 599–610, ACM, 2013.

[21] A. Serwadda, V. V. Phoha, Z. Wang, R. Kumar, and D. Shukla, "Toward robotic robbery on the touch screen," *ACM Transactions on Information and System Security*, vol. 18, no. 4, p. 14, 2016.

[22] R. Kumar, V. V. Phoha, and A. Jain, "Treadmill attack on gait-based authentication systems," in *IEEE 7th International Conference on Biometrics Theory, Applications and Systems*, pp. 1–7, IEEE, 2015.

[23] A. K. Jain, A. Ross, and K. Nandakumar, *Introduction to Biometrics*. Springer, 2011.

[24] M. Barcena, C. Wueest, and H. Lau, "How safe is your quantified self?," *Symantec*, pp. 1–38, 2014.

[25] J. Rieck, "Attacks on fitness trackers revisited: A case-study of unfit firmware security," *Lecture Notes in Informatics*, 2016.

[26] J. Benaloh, M. Chase, E. Horvitz, and K. Lauter, "Patient controlled encryption," in *Proceedings of the 2009 ACM workshop on Cloud Computing Security*, p. 103, ACM, 2009.

[27] S. Eberz, K. B. Rasmussen, V. Lenders, and I. Martinovic, "Preventing lunchtime attacks: Fighting insider threats with eye movement biometrics," in *22th Annual Network and Distributed System Security Symposium*, 2015.

[28] D. Jang, R. Jhala, S. Lerner, and H. Shacham, "An empirical study of privacy-violating information flows in javascript web applications," in *Proceedings of the 17th ACM conference on Computer and Communications Security*, p. 270, ACM, 2010.

[29] A. Kapravelos, C. Grier, N. Chachra, C. Kruegel, G. Vigna, and V. Paxson, "Hulk: Eliciting malicious behavior in browser extensions.," in *USENIX Security Symposium*, pp. 641–654, 2014.

[30] A. Weiss, A. Ramapanicker, P. Shah, S. Noble, and L. Immohr, "Mouse movements biometric identification: A feasibility study mouse movement biometric system," *Proceedings of Student Faculty Research Day CSIS Pace University*, pp. 1–8, 2007.

[31] J. Hohl and S. Rush, "The complete heart-lead relationship in the einthoven triangle," *The Bulletin of Mathematical Biophysics*, vol. 30, no. 4, pp. 615–623, 1968.

[32] N. Zheng, A. Paloski, and H. Wang, "An efficient user verification system via mouse movements," in *Proceedings of the 18th ACM conference on Computer and Communications Security*, p. 139, ACM, 2011.

[33] T. G. Kolda, R. M. Lewis, and V. Torczon, "Optimization by direct search: New perspectives on some classical and modern methods," *SIAM review*, vol. 45, no. 3, pp. 385–482, 2003.

[34] B. C. Ross, A. Kraskov, H. Stögbauer, P. Grassberger, I. Grosse, P. Bernaola-Galván, P. Carpena, R. Román-Roldán, J. Oliver, L. Kozachenko, and N. Leonenko, "Mutual information between discrete and continuous data sets," *PLoS ONE*, vol. 9, no. 2, pp. 1–5, 2014.

Appendix A. Complete Feature Analysis

In Table 5, 6, 7, 8, 9, we report the results of the computation of the unpredictability scores (Equation 5) for individual features, for ECG, gait, touch dynamics, mouse movements and eye movements, respectively. The tables also report the complete list of features used in the paper and their RMI (with the exception for gait that only shows the results for the 25 most relevant features, for brevity). We used the methodology described in Section 5.3 and the target contexts explained in Section 6.1.

Appendix B. Optimization Problem

Let $D_{f_\theta} = \left\{\xi \in \mathbb{R} | F_{f_\theta(X_u^{(s)})}(\xi) - F_{X_u^{(t)}}(\xi) \neq 0\right\}$ be the set in which the integrand of Equation 3 is non zero. In general this depends upon the specific choice of the mapping function f_θ. Let $lb^{(t)}$ be the minimum observed value for the target featureand $ub^{(t)}$ be the maximum, then, we define a region around the observed target feature space as:

$$D = \left[lb^{(t)} - q\Delta^{(t)}, ub^{(t)} + q\Delta^{(t)}\right]$$

where $\Delta^{(t)} = ub^{(t)} - lb^{(t)}$ and $q > 0$ is a relaxing factor. Given a parametrized family of mapping functions $\{f_\theta\}_{\theta \in \Theta}$, we then define Θ to be:

$$\Theta = \{\theta \in \mathbb{R}^m | D_{f_\theta} \subseteq D\}\,. \tag{7}$$

This constrains the range of the transformed source feature in a way that it is similar to the range of the target feature. Additionally, for every $\xi \in \mathbb{R}$ we have that $0 \leq (F_{f_\theta(X_u^{(s)})}(\xi) - F_{X_u^{(t)}}(\xi))^2 \leq 1$, thus for every $\theta \in \Theta$, we have that:

$$0 \leq \int_D \left(F_{f_\theta(X_u^{(s)})}(\xi) - F_{X_u^{(t)}}(\xi)\right)^2 d\xi \leq m(D)$$

where $m(D)$ is the measure of D. Thus, by setting $w = 1/m(D)$, we have that the distance function defined in Equation 3 is normalized in $[0, 1]$.

Finally, notice that the set of constraints 7 is linear if f_θ is linear with respect to the parameter vector θ (e.g. f_θ is a generic polynomial function) and hence can be solved exactly (and efficiently) by an optimization algorithm. In fact, in this case we have that $f_\theta(x) = \theta^T \cdot \mathbf{h}(x)$, for a specific vector valued function $\mathbf{h}$. Hence constraint 7 is equivalent to checking for each feature value x observed in the users population:

$$lb^{(t)} - q\Delta^{(t)} \leq \theta^T \cdot \mathbf{h}(x) \leq ub^{(t)} + q\Delta^{(t)}$$

which is linear in the parameter vector θ.

Appendix C. Degree Analysis

We list in Table 10 biometric unpredictability scores for cross-context mappings of degrees from 1 to 4, using source and target context pairs discussed in Section 6.1. Figure 7 shows how unpredictability scores are affected by different polynomials f_θ in the case of ECG biometric. Figure 7 and Table 10 show that the relative ranking in terms of average unpredictability scores for the biometrics is not affected by the degree of f_θ. For the 41 different source context pairs that we have discussed a linear f_θ provides an estimation of the unpredictability score similar to that of polynomials f_θ.

The score of mouse movement biometrics is the only one that slightly decreases as the degree of f_θ increases,

ECG	RMI(%)	*Lead I-rest*	*Lead II-rest*	*Lead III-rest*	*Palm-rest*	*mobile-rest*	*Fitness tracker-rest*	*Fitness tracker-walk*	*Fitness tracker-jog*
SAmplitude	44.3	.05 ± .02	.08 ± .03	.09 ± .04	.05 ± .02	.04 ± .02	.05 ± .02	.05 ± .02	.06 ± .01
TAmplitude	44.1	.11 ± .03	.12 ± .03	.09 ± .02	.06 ± .01	.04 ± .01	.09 ± .02	.09 ± .02	.10 ± .02
STInterval	33.4	.08 ± .02	.11 ± .03	.15 ± .03	.08 ± .02	.09 ± .02	.11 ± .03	.14 ± .02	.15 ± .03
SDuration	32.7	.07 ± .03	.09 ± .02	.10 ± .02	.07 ± .01	.05 ± .02	.08 ± .02	.08 ± .02	.07 ± .02
PAmplitude	31.3	.05 ± .01	.15 ± .03	.15 ± .03	.06 ± .02	.09 ± .02	.15 ± .03	.14 ± .04	.12 ± .04
SDuration_R	28.8	.07 ± .03	.09 ± .03	.09 ± .02	.06 ± .01	.05 ± .02	.07 ± .01	.09 ± .02	.08 ± .02
RDuration	27.1	.07 ± .01	.10 ± .03	.13 ± .03	.07 ± .01	.06 ± .01	.13 ± .03	.11 ± .03	.12 ± .03
PQInterval	25.6	.06 ± .01	.09 ± .02	.11 ± .03	.13 ± .03	.07 ± .01	.10 ± .03	.12 ± .03	.13 ± .03
SDuration_L	24.7	.08 ± .02	.10 ± .03	.11 ± .02	.07 ± .01	.07 ± .02	.10 ± .03	.10 ± .03	.11 ± .03
RSInterval	23.4	.09 ± .03	.16 ± .06	.14 ± .05	.10 ± .02	.07 ± .02	.17 ± .06	.14 ± .05	.15 ± .06
RDuration_R	21.0	.07 ± .01	.11 ± .03	.13 ± .03	.06 ± .01	.07 ± .01	.14 ± .03	.15 ± .03	.13 ± .03
QDuration	20.6	.09 ± .02	.12 ± .03	.11 ± .03	.10 ± .03	.14 ± .04	.11 ± .02	.11 ± .03	.12 ± .03
QDuration_L	20.2	.08 ± .02	.11 ± .03	.11 ± .03	.10 ± .03	.13 ± .05	.10 ± .02	.10 ± .03	.11 ± .02
QAmplitude	20.0	.05 ± .02	.04 ± .02	.06 ± .03	.08 ± .01	.04 ± .02	.04 ± .01	.02 ± .01	.04 ± .01
TDuration	18.9	.05 ± .02	.07 ± .02	.09 ± .02	.08 ± .01	.05 ± .01	.11 ± .02	.08 ± .02	.09 ± .02
QDuration_R	18.1	.12 ± .02	.11 ± .02	.11 ± .02	.10 ± .02	.13 ± .02	.11 ± .02	.10 ± .03	.09 ± .02
QRInterval	17.8	.14 ± .02	.19 ± .04	.22 ± .04	.18 ± .03	.20 ± .03	.18 ± .04	.16 ± .05	.17 ± .04
RDuration_L	17.2	.06 ± .01	.08 ± .02	.11 ± .02	.07 ± .01	.06 ± .01	.11 ± .02	.12 ± .03	.09 ± .02
TDuration_R	14.0	.04 ± .01	.07 ± .02	.07 ± .01	.05 ± .01	.05 ± .01	.07 ± .02	.05 ± .01	.05 ± .01
TDuration_L	13.1	.07 ± .02	.09 ± .03	.13 ± .02	.08 ± .02	.06 ± .02	.12 ± .02	.10 ± .02	.10 ± .02
PDuration	9.9	.09 ± .02	.13 ± .03	.15 ± .03	.11 ± .02	.10 ± .02	.15 ± .03	.11 ± .03	.12 ± .03
PDuration_R	8.4	.08 ± .02	.08 ± .02	.11 ± .02	.10 ± .02	.08 ± .02	.11 ± .02	.10 ± .02	.08 ± .02
PDuration_L	5.4	.07 ± .02	.09 ± .02	.10 ± .02	.08 ± .01	.09 ± .02	.09 ± .02	.08 ± .02	.08 ± .02

TABLE 5: Complete RMI values and unpredictability scores for each feature in the ECG biometric, computed on the first session. The confidence intervals are computed over the result for each user.

though the difference between the linear and the degree 4 unpredictability scores is still not significant as for a one-tailed Wilcoxon rank sum tests. This happens as the aggregated distribution of the $click\_duration$ feature (the most relevant for the mouse movements biometric) for the trackpad has several modes depending on whether the user uses the trackpad click button or simply taps it, whereas the $click\_duration$ for mouse device is much more consistent across users. Hence, using a quadratic polynomial provides an advantage over linear function, however this quickly saturates for degrees greater than 3.

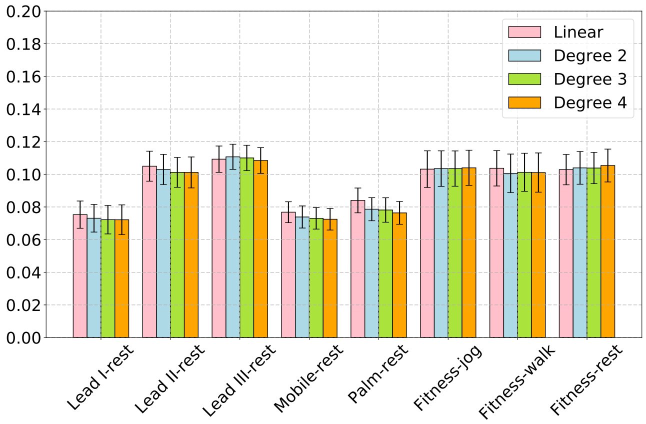


Figure 7: Unpredictability score as estimated by polynomials of degree from 1 (linear) to 4, for ECG biometric. Errorbars show 95% confidence intervals.

Gait	RMI(%)	*Smartwatch-walk*	*Hand phone-walk*	*Smartwatch-jog*	*Cheststrap-jog*	*Arm phone-jog*	*Pocket phone-jog*	*Arm phone-walk*	*Cheststrap-walk*
f_36	19.4	.18 ± .04	.18 ± .04	.18 ± .04	.16 ± .03	.15 ± .04	.11 ± .03	.16 ± .04	.18 ± .03
f_42	19.3	.21 ± .04	.15 ± .04	.14 ± .04	.18 ± .04	.18 ± .04	.15 ± .03	.15 ± .03	.16 ± .03
f_43	19.3	.21 ± .04	.16 ± .04	.15 ± .04	.17 ± .04	.15 ± .04	.17 ± .04	.16 ± .03	.18 ± .04
f_41	19.1	.18 ± .04	.14 ± .03	.13 ± .03	.17 ± .03	.17 ± .04	.14 ± .03	.14 ± .03	.16 ± .03
f_7	19.0	.17 ± .04	.18 ± .04	.15 ± .04	.18 ± .04	.17 ± .05	.12 ± .04	.16 ± .04	.16 ± .03
f_8	18.8	.16 ± .04	.16 ± .04	.17 ± .04	.19 ± .04	.16 ± .05	.12 ± .04	.15 ± .04	.18 ± .04
f_35	18.8	.19 ± .04	.18 ± .04	.18 ± .04	.16 ± .03	.15 ± .03	.11 ± .03	.17 ± .04	.19 ± .03
f_6	18.8	.18 ± .04	.18 ± .04	.15 ± .04	.17 ± .04	.17 ± .04	.12 ± .03	.16 ± .03	.17 ± .03
f_4	18.7	.17 ± .03	.15 ± .04	.18 ± .04	.14 ± .03	.17 ± .04	.13 ± .03	.14 ± .03	.16 ± .03
f_64	18.7	.14 ± .03	.14 ± .03	.15 ± .04	.15 ± .03	.17 ± .04	.13 ± .02	.15 ± .04	.17 ± .04
f_63	18.7	.12 ± .03	.13 ± .04	.14 ± .04	.14 ± .03	.16 ± .04	.12 ± .03	.14 ± .03	.15 ± .04
f_5	18.6	.18 ± .04	.18 ± .04	.17 ± .05	.15 ± .03	.19 ± .04	.13 ± .03	.16 ± .03	.16 ± .03
f_44	18.5	.19 ± .04	.18 ± .03	.16 ± .04	.17 ± .04	.15 ± .04	.15 ± .03	.16 ± .03	.18 ± .03
f_0	18.4	.14 ± .03	.15 ± .03	.16 ± .03	.19 ± .02	.19 ± .03	.11 ± .02	.13 ± .03	.18 ± .03
f_28	18.4	.18 ± .05	.17 ± .04	.17 ± .04	.18 ± .04	.18 ± .05	.13 ± .04	.16 ± .03	.19 ± .04
f_69	18.4	.16 ± .03	.16 ± .03	.13 ± .02	.16 ± .03	.15 ± .03	.12 ± .02	.13 ± .04	.17 ± .03
f_17	18.2	.14 ± .03	.17 ± .04	.16 ± .04	.14 ± .03	.17 ± .04	.14 ± .03	.15 ± .04	.14 ± .04
f_1	18.1	.14 ± .03	.15 ± .03	.17 ± .02	.17 ± .03	.17 ± .03	.12 ± .03	.13 ± .04	.18 ± .04
f_37	18.1	.19 ± .03	.17 ± .03	.18 ± .03	.16 ± .03	.14 ± .03	.11 ± .03	.15 ± .04	.16 ± .03
f_3	18.0	.16 ± .03	.14 ± .03	.17 ± .04	.14 ± .03	.16 ± .04	.12 ± .03	.13 ± .03	.15 ± .03
f_29	17.8	.19 ± .04	.16 ± .03	.17 ± .03	.16 ± .03	.17 ± .04	.12 ± .03	.15 ± .03	.16 ± .03
f_16	17.6	.15 ± .03	.18 ± .04	.15 ± .04	.13 ± .04	.17 ± .04	.14 ± .03	.16 ± .04	.16 ± .04
f_65	17.6	.17 ± .04	.12 ± .03	.14 ± .03	.15 ± .03	.18 ± .04	.11 ± .02	.15 ± .03	.17 ± .04
f_9	17.4	.14 ± .03	.17 ± .04	.17 ± .04	.18 ± .04	.17 ± .04	.13 ± .04	.14 ± .04	.17 ± .03
f_40	17.4	.17 ± .04	.15 ± .03	.12 ± .03	.15 ± .03	.18 ± .04	.12 ± .03	.13 ± .03	.13 ± .03

TABLE 6: Complete RMI values and unpredictability scores for each feature in the gait biometric, computed on the first session. The confidence intervals are computed over the result for each user.

Touch dynamics	RMI(%)	*Low-end phone*	*High-end phone*
mid_stroke_pressure	16.3	.15 ± .03	.15 ± .03
mid_stroke_area_covered	16.0	.15 ± .03	.11 ± .03
stroke_duration	15.8	.04 ± .01	.04 ± .01
y_end	14.2	.09 ± .02	.09 ± .02
direct_end_to_end_distance	13.7	.09 ± .02	.07 ± .02
length_of_trajectory	13.5	.07 ± .02	.06 ± .01
direction_of_end_to_end_line	12.1	.05 ± .02	.04 ± .01
y_start	11.9	.09 ± .02	.09 ± .02
x_start	11.5	.06 ± .01	.05 ± .01
x_end	10.3	.07 ± .02	.06 ± .01
f_20_perc_perp_distance	8.5	.05 ± .01	.03 ± .01
f_50_perc_perp_distance	8.4	.06 ± .01	.03 ± .01
f_80_perc_perp_distance	7.7	.05 ± .01	.03 ± .01
max_perp_distance	7.5	.06 ± .01	.03 ± .01
median_accel_first5	7.5	.15 ± .02	.07 ± .01
average_velocity	6.8	.07 ± .02	.08 ± .02
average_direction	6.7	.05 ± .01	.05 ± .01
median_velocity_last5	6.0	.08 ± .02	.10 ± .02
f_50_perc_accel	5.3	.10 ± .02	.09 ± .02
f_80_perc_velocity	4.9	.08 ± .02	.09 ± .02
f_20_perc_accel	4.5	.07 ± .01	.08 ± .01
f_50_perc_velocity	4.2	.09 ± .02	.08 ± .02
f_80_perc_accel	3.6	.13 ± .02	.07 ± .01
mean_resultant_length	3.0	.05 ± .01	.05 ± .01
f_20_perc_velocity	2.6	.14 ± .01	.07 ± .01
distance_to_traj_length	2.4	.04 ± .01	.04 ± .01
median_velocity_last3	2.0	.08 ± .02	.07 ± .02

TABLE 7: Complete RMI values and unpredictability scores for each feature in the touch dynamics biometric, computed on the first session. The confidence intervals are computed over the result for each user.

Mouse movements	RMI(%)	*Trackpad*
click_duration	15.7	.12 ± .03
curvature_distances_mean	4.1	.04 ± .01
curvature	3.3	.03 ± .01
curvature_distances_max	3.2	.02 ± .01
angle_of_curvature_max	2.9	.05 ± .01
curvature_distances_std	2.6	.03 ± .01
acc_dev	2.4	.06 ± .02
angle_of_curvature_mean	2.4	.05 ± .02
speed_mean	2.2	.06 ± .02
straight_dev_std	2.0	.04 ± .01
angle_of_curvature_std	2.0	.07 ± .02
speed_max	1.9	.06 ± .01
straight_dev_max	1.8	.06 ± .02
acc_max	1.8	.04 ± .01
speed_dev	1.8	.07 ± .02
acc_mean	1.6	.03 ± .01
straight_dev_mean	1.5	.04 ± .01
acc_min	0.8	.04 ± .01
curvature_distances_min	0.7	.02 ± .00
angle_of_curvature_min	0.5	.02 ± .01

TABLE 8: Complete RMI values and unpredictability scores for each feature in the mouse movements biometric, computed on the first session. The confidence intervals are computed over the result for each user.

Eye movements	RMI(%)	*Intra task-uncalibrated*	*Cross task-uncalibrated*
pupil_max	19.5	.03 ± .01	.04 ± .01
pupil_mean	19.4	.03 ± .01	.04 ± .01
pupil_min	18.8	.06 ± .01	.06 ± .02
pupil_range	4.5	.03 ± .01	.04 ± .01
accel_mean	4.3	.03 ± .01	.04 ± .01
pupil_std	3.8	.06 ± .02	.07 ± .02
speed_mean	3.8	.03 ± .01	.03 ± .01
fix_duration	2.7	.03 ± .01	.04 ± .01
centredist_mean	1.8	.04 ± .01	.05 ± .01
accel_max	1.2	.02 ± .01	.03 ± .01
speed_std	1.2	.03 ± .01	.03 ± .01
speed_max	0.8	.04 ± .01	.05 ± .01
centredist_max	0.7	.05 ± .01	.05 ± .01
centredist_min	0.6	.08 ± .02	.10 ± .02
maxpwdist_y	0.5	.08 ± .02	.10 ± .02
maxpwdist_x	0.5	.08 ± .02	.10 ± .02
centredist_std	0.3	.04 ± .01	.05 ± .01
maxpwdist	0.3	.03 ± .01	.04 ± .01

TABLE 9: Complete RMI values and unpredictability scores for each feature in the eye movements biometric, computed on the first session.The confidence intervals are computed over the result for each user.

	Linear	Degree 2	Degree 3	Degree 4
ECG	.095 ± .004	.093 ± .004	.092 ± .004	.092 ± .004
Eye	.076 ± .014	.078 ± .018	.079 ± .014	.079 ± .014
Mouse	.068 ± .010	.062 ± .009	.058 ± .008	.058 ± .008
Touch	.077 ± .006	.077 ± .006	.076 ± .006	.076 ± .006
Gait	.153 ± .008	.153 ± .008	.152 ± .008	.152 ± .008

TABLE 10: Average unpredictability score for each biometric, for cross-context functions of degrees from 1 to 4. For each source we also show the 95% confidence intervals computed over the unpredictability scores of individual users.