Improving Accuracy, Applicability and Usability of Keystroke Biometrics on Mobile Touchscreen Devices

Daniel Buschek[1], **Alexander De Luca**[1,2], **Florian Alt**[1]
[1]Media Informatics Group, University of Munich (LMU); [2]DFKI GmbH, Saarbrücken, Germany
{daniel.buschek, alexander.de.luca, florian.alt}@ifi.lmu.de

ABSTRACT

Authentication methods can be improved by considering implicit, individual behavioural cues. In particular, verifying users based on typing behaviour has been widely studied with physical keyboards. On mobile touchscreens, the same concepts have been applied with little adaptations so far. This paper presents the first reported study on mobile keystroke biometrics which compares touch-specific features between three different hand postures and evaluation schemes. Based on 20.160 password entries from a study with 28 participants over two weeks, we show that including spatial touch features reduces implicit authentication equal error rates (EER) by 26.4 - 36.8% relative to the previously used temporal features. We also show that authentication works better for some hand postures than others. To improve applicability and usability, we further quantify the influence of common evaluation assumptions: known attacker data, training and testing on data from a single typing session, and fixed hand postures. We show that these practices can lead to overly optimistic evaluations. In consequence, we describe evaluation recommendations, a probabilistic framework to handle unknown hand postures, and ideas for further improvements.

Author Keywords

Keystroke Dynamics; Mobile; Touch; Biometrics

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation (e.g. HCI): Input devices and strategies (e.g. mouse, touchscreen)

INTRODUCTION

We use mobile devices in many tasks every day [8]. Some require access to password-protected systems, like email accounts or social networks. The device itself may also be locked with a password or PIN to protect it in cases of loss or theft [25]. Utilising device sensors and assuming a one-to-one relationship with personal devices, research proposed behavioural biometrics to enhance security [11, 16, 18, 31]. For password entry, a second, implicit security layer can observe typing *behaviour*: If an attacker knows the password or PIN, for example due to shoulder surfing [48] or a smudge-attack [2, 51], access can still be denied based on the fact that they do not type the password in the same way as the legitimate user (e.g. different rhythm, finger placement).

CHI 2015, April 18 - 23 2015, Seoul, Republic of Korea

ACM 978-1-4503-3145-6/15/04...$15.00
http://dx.doi.org/10.1145/2702123.2702252

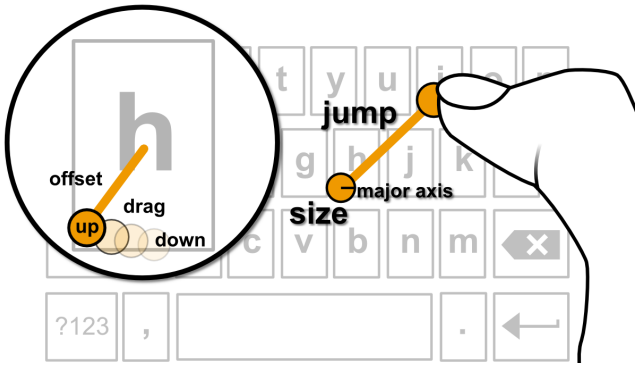

Figure 1. Touch-specific keystroke features on a mobile keyboard. In this example, the user is typing "hi". The magnified *h*-key shows touch down and up locations, the drag in between, and the offset to the key's centre. The keyboard-overlay shows touch area size and axis, as well as the "jump" vector between subsequent touches. In this paper, we analyse these touch-specific features to improve mobile keystroke biometrics.

Verifying identity based on typing behaviour (*keystroke dynamics*) has mostly been studied in terms of timing, both on physical keyboards [40, 41] and older mobile devices with physical keys [10, 14, 15, 30, 33, 38, 57]. Less work has been carried out on mobile devices with touch [32, 42, 46, 59], and without investigating touch-specific behavioural features, such as those shown in Figure 1. Recent research has successfully used similar features with gesture keyboards [11], but the potential of touch biometrics is still unknown for typing by tapping [20, 55]. Hence, to improve keystroke biometrics for the smartphone-era, this paper provides in-depth analyses of mobile-specific and touch-specific opportunities and challenges, leading to the following contributions:

1. *To improve implicit authentication accuracy, we evaluate touch-specific features for capturing individual typing behaviour.* Spatial touch features outperform the commonly used temporal features, and both can be combined to reduce equal error rates by up to 36.8%.
2. *To improve applicability, we discuss and quantify practical implications of different commonly used evaluations.* In particular, we compare results for: 1) training and testing within sessions or across sessions; 2) training on owner data only or also on data from others; 3) assuming fixed or changing hand postures. Our analyses allow for a more realistic assessment of keystroke biometrics in practice.
3. *To improve usability, we propose an approach to avoid restricting users to one typing posture.* We analyse one-thumb, two-thumb and index finger typing. We show that behaviour is highly posture-specific and present a method to handle changing hand postures.

RELATED WORK

We relate our work to touch behaviour, keystroke dynamics, and touch-based implicit identification and authentication.

Modelling Touch and Typing Behaviour

Related research reduced typing errors with keyboard personalisation based on users' individual touch distributions per key [3, 4, 13, 21, 24, 56]. Azenkot and Zhai [3] found that touch distributions for each key on a smartphone's soft keyboard varied less within users than between them. Yin et al. [56] showed that user-specific models reduce error rates further than posture- or key-specific ones. Edelmann et al. [21] also found that user-specific models led to less typing errors on a tabletop. Hence, this line of research revealed individual touch typing patterns, but none of these projects utilised this information for user authentication, which we explore here.

Research also found several influences on general touch behaviour: Individuals visually target differently with their finger tips [27], since targets are often occluded by the finger [6]. Contact areas are influenced by the finger's angle [22, 52]. Targeting offsets differ with respect to the finger's pitch, roll and yaw [28], and offset patterns [26] can better be corrected for individual users than overall [12, 53, 54]. In this paper, we examine targeting behaviour in terms of such touch offsets as a novel feature for mobile keystroke biometrics.

Behavioural Biometrics for Mobile Typing

Related work applied keystroke dynamics on mobile phones with physical keys: Zahid et al. [57] deployed a fuzzy-logic classifier observing temporal features and error counts on a keypad phone. Other researchers used neural networks to authenticate mobile phone users based on temporal typing features on keypads [15] and physical mini-QWERTY keyboards [33]. Their systems achieved 12.8% EER [15] and 12.2% EER [33], using attacker data during training.

Nauman et al. [42] enhanced password authentication for web services with keystroke dynamics on smartphones. They used keystroke latencies and key-hold times, but no touch features. Saevanee et al. [47] combined linguistic analysis with keystroke dynamics on mobile devices. The keystroke-based part of their system used key-hold times and achieved 20.8% EER. They trained classifiers in a "one-vs-rest" scheme, using data from both the legitimate user and others.

Zheng et al. [59] measured pressure and touch size for nine-key PIN unlock on smartphones. They reported EERs between 3.65%-7.34% with distance-based anomaly detection evaluated in a single session. Other authors used a Bayesian Net to authenticate phone users in a passcode entry task with a grid of 16 symbols [32]. They reported 82.18% accuracy with temporal features and distances between subsequent touches. They also applied the method to an actual keyboard, but did not consider touch features there. In contrast to these projects, we include touch offsets and touch locations.

Burgbacher and Hinrichs [11] trained Support Vector Machines (SVMs) in a "one-vs-rest" scheme to authenticate users via finger movement behaviour on gesture keyboards (see [36]). They reported 0% EER if five or more words in the message are known to the system. They targeted gesture keyboards, whereas our approach addresses typing by tapping.

Draffin et al. [20] trained neural network classifiers for keystroke authentication using touch-to-key offsets, size, pressure, drag, and hold time. Although they used features derived from exact touch locations, they did not evaluate the contribution of these new spatial features to the overall performance. This was also pointed out by Xu et al. [55], who therefore decided against using these typing features in their work on implicit authentication with mobile touch input. In contrast, we present a detailed evaluation of offsets and other touch features. We show that they outperform the temporal features, and thus should indeed be included.

Touch-based Implicit Authentication and Identification

Related work addressed verifying user identity with diverse touch measures: Shape-based phone unlock systems were enhanced with an implicit layer using touch sequence matching [1, 7, 18]. Characteristics of touch strokes from zooming and scrolling were utilised as well [23, 58]. Other work suggested to directly replace passwords with touch evidence [31, 49], for example with special touch gestures [45]. In contrast, our method addresses text-based logins and typing.

Further related research distinguished users with rear-projected tabletop systems: Holz and Baudisch [29] used a fiber optics plate to authenticate users via fingerprints during touch interaction on a multitouch table. Mock et al. [39] also captured images of finger contact areas on a tabletop. They used SVMs to identify one of twelve known typists with 97.51% accuracy, and detected unknown users with 12.3% EER, both after one keystroke. This shows that finger placement on soft keyboards provides user-specific information. However, current mobile touchscreens lack optical sensors; images of the fingers are not available. In this paper, we nevertheless utilise spatial typing touch information, solely relying on sensors available on off-the-shelf mobile devices.

Opportunities and Intended Contribution

In summary, related work on keyboard personalisation, targeting, and touch-based authentication has shown individual touch and typing behaviour. However, research on keystroke biometrics has either ignored spatial touch-specific typing features on mobile devices [15, 33, 42, 57], or only used such features on tabletops [29, 39] or with gesture-keyboards [11].

We found only one exception: Draffin et al. [20] measured touch-to-key offsets, but only tested them with one classifier, and crucially neither optimised feature sets nor evaluated the influence of the new features, as pointed out by Xu et al. [55].

In conclusion, this leaves the potential of touch features for mobile keystroke biometrics still unexplored. At the same time, related work has revealed the need to address mobile applications of keystroke biometrics and to develop novel features [5, 16, 50]. Hence, we aim to improve behavioural biometrics for mobile touch keyboards with a comprehensive analysis of touch-specific features in a password entry task.

Furthermore, we analyse common evaluation schemes in research on (mobile) keystroke biometrics to quantify the effects of potentially "optimistic" methods. We suggest alternatives which allow for more practical and usable perspectives.

THREAT MODEL

We consider that an attacker gains access to an unlocked device and additionally knows the owner's password for their email account, for a social network or for similar services and apps which require authentication. Here, keystroke information serves as an additional security layer ("password hardening" [40]): even if the attacker enters the correct password, the system can deny access due to different typing behaviour.

IMPROVING ACCURACY OF KEYSTROKE BIOMETRICS

We propose new typing features for the password entry task resulting from the described threat model. We then introduce our employed user models for implicit authentication.

Typing Features

To build user models for authentication, we need to capture individual aspects of typing behaviour. We are interested in features which vary characteristically between users, but also stay consistent for the same user over time. Formally, we describe each password entry as a feature vector, as in most related work (see e.g. [20, 42, 50]): Typing a password with n keystrokes is represented as a vector $f = (f_1, f_2, ..., f_n)^T$, the concatenation of the typing features f_t for each touch t. For example, if a system observes the two features *hold time* ht and *flight time* ft, the typing behaviour for a password entry is described as $f = (ht_1, ft_1, ht_2, ft_2, ..., ht_n, ft_n)^T$. This representation of user behaviour can then be processed by machine learning methods to authenticate users.

Related work on keystroke-based biometrics commonly uses temporal typing features (e.g. [15, 33, 42, 57]): *hold time* passes between the moment when the finger touches the screen (touch down), and the moment when the finger is lifted (touch up). Complementary, *flight time* is measured between touch up and touch down. We can also measure the time in between subsequent touch up or down events, to which we refer as *up-up* times and *down-down* times, respectively.

To complement the temporal features, we propose to consider new spatial touch-specific features (Figure 1): *exact touch locations* at touch down and up events; *offsets* between touch up and key-centres, due to typing inaccuracy; touch *"jumps"*, the distances between subsequent touches; *drag* distances/angles between touch down and up locations, due to small "natural" movements ($\approx$ 2.1 pixels average drag in our data); touch area *sizes* and ellipses *axes*. We also evaluate touch *pressure*. Size, axes and pressure are estimated by the Android API.

User Models for Authentication

We compare models of two types: 1) anomaly detectors, which only require training data from the legitimate user; and 2) classification methods, which are trained on data from multiple users. In practice, training data can be collected in an enrolment phase or from normal use. During testing, that means whenever a password is entered correctly, these models then decide whether it was typed by the legitimate user.

Anomaly Detection: Many models for keystroke-based authentication exist [5, 34, 50]. To show that results are not model-specific, we selected methods for three common approaches: distance to training instances (here: mean distance to k=5 nearest neighbours, *kNN*); a statistical model (here: Gaussian model without covariance, *GM*); and a kernel-based method (here: Least Squares Anomaly Detection, *LSAD* [44]). We refer to the related work for detailed descriptions.

Classification: Complementary to authentication, we chose a small set of representative classifiers: k-Nearest-Neighbour classification (*kNNC*); Naïve Bayes (*NB*); and Support Vector Machines (*SVM*). We refer to related work for more details on these models [19, 43] and their applications [5, 50].

IMPROVING APPLICABILITY AND USABILITY

To improve mobile keystroke biometrics besides raw accuracy, we further target applicability and usability. Our goal is to 1) identify evaluations and assumptions that can result in too optimistic assessments of the actually applicable and usable quality of the examined systems; and to 2) propose and highlight more usable and practically relevant alternatives.

In particular, we compare conditions and quantify effects of: collecting data in a single session - or in multiple sessions; training user models on data from the owner - or also on data from others; and assuming a fixed known hand posture - or leaving the choice of postures to the user. The following sections discuss these issues in more detail and introduce our proposed improved concepts for evaluation and applications.

Evaluation Within Sessions vs Across Sessions

Typing and touch behaviour can be expected to vary over time. Therefore, training and testing on data from the same session is likely to be too optimistic, since, in practical applications, enrolment and authentication will never follow directly one after the other. However, a recent survey [50] found that 73% of examined publications on keystroke behavioural biometrics only studied data obtained in a single session.

To improve mobile keystroke biometrics for practical use, it is thus important to study the practically relevant case across sessions, and to quantify the effects of single session evaluation to inform future study design. To the best of our knowledge, this is the first reported direct comparison of evaluations within and across sessions on mobile touch keystroke data. We show that single session evaluations can lead to EERs less than half of those obtained across sessions. We thus recommend to collect data in at least two sessions.

Classification vs Anomaly Detection Methods

Two training and modelling schemes are commonly employed for keystroke-based biometrics: classification and anomaly detection (Figure 2). Classifiers are potentially more powerful, since they characterise the legitimate user in *contrast* to others, whereas anomaly detectors can only check for deviation from the legitimate user's behaviour. However, the need for data from others can make classifiers difficult or even impossible to apply in practice.

Since classifiers need training data from multiple users, their application to capture behaviour for specific passwords or PINs implies that the user's password or PIN has been typed by others and is thus known to them. In contrast to the common use of classifiers in this way [15, 32, 42, 46], we thus strongly recommend to focus on anomaly detectors instead, when adding an implicit layer for password or PIN entry.

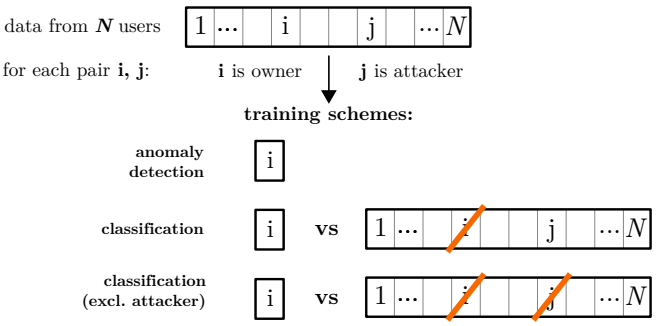

Figure 2. Three *training schemes for evaluation* of keystroke biometrics. Anomaly detectors are only trained on data from the legitimate user ("owner"). In contrast, classifiers also use pooled data from all other users. However, applications may not always have access to typing behaviour of other users in practice, especially not for specific secret passwords. Moreover, it is unrealistic to assume known data from the attacker. To address these issues, we 1) recommend anomaly detection for applications where features are extracted for secret passwords, and we 2) propose a slightly different training scheme for classifiers, which excludes the attacker from the training data for the "others"-class.

Moreover, the dataset of others used to train classifiers often also includes data from the very participant who serves as an attacker in the current evaluation case [17]. Assuming that data from the attacker is known to the system in advance is highly unrealistic.

We address this issue with a simple alternative way of splitting the training data for evaluation of classifiers (Figure 2). In particular, we split the data into *three* parts: owner, attacker, and others (excluding the attacker). Classifiers can then be trained on data from owner and others, without assuming known data from the attacking individual.

Note that we do not argue against the use of classifiers for keystroke biometrics altogether: A system could be shipped with anonymous typing data for common words (but not passwords or PINs), collected by the developers in a user study. This data can then be used to train models against the device owner's data for the same words, observed when typing messages, e-mails, and so on (see e.g. [11, 59]). Finally, classifiers trained on "shipped data vs owner data" in this way may also be applied (even to passwords) if typing behaviour is not described per word but for example per bigram [50]. These cases are not further addressed in our study in this paper.

To the best of our knowledge, we are the first to report a direct comparison of anomaly detectors and classifiers on mobile touchscreen keystroke data. We show that classifiers lead to 28.4 - 48.1% lower EERs relative to anomaly detectors, and to 45.2 - 58.2% lower EERs, if the attacker's data is included. We thus recommend to carefully consider if classifiers are practically applicable for given use-cases. We further recommend to exclude the attackers' data from training sets when evaluating such systems.

Fixed vs Changing Hand Postures

Many studies of mobile keystroke biometrics evaluated systems with data from one hand posture or did not report the posture [14, 15, 32, 33, 57]. Some reported results for different postures [30, 59], but those were always treated separately. As a result for practical deployments, these evaluations imply systems that require fixed hand postures: the user would always have to type with the same posture for the system to work with the evaluated accuracy. This clearly restricts the user's freedom and lowers the usability of such a system.

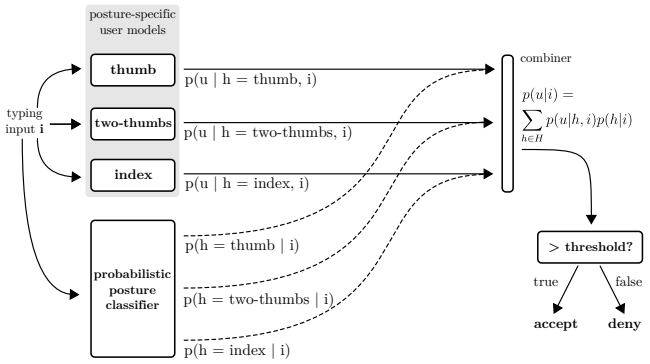

Figure 3. Probabilistic framework for usable mobile keystroke biometrics which does not restrict users to a fixed hand posture. For an entered password (input i), each posture-specific model predicts the probability $p(u|h,i)$ of the legitimate user u, assuming that the corresponding hand posture h was used. Additionally, a posture classifier estimates the probability $p(h|i)$ of each posture indeed being the one used while typing. Both sets of probabilities are then combined. The resulting probability $p(u|i)$ of the legitimate user can then be compared against a threshold.

To the best of our knowledge, this paper presents the first reported direct comparison of mobile touch keystroke biometrics for different postures. We show that entering a password in a system trained on a different posture increases EERs by up to 86.3% relative to a system assuming a fixed posture.

To address this issue and improve usability of keystroke biometrics, we propose a framework that allows users to type with different postures. We follow a probabilistic approach, utilising the fact that we do not need to *decide* for a certain posture, since we only care about correct authentication.

In summary, when a password is entered, we use a probabilistic classifier to predict a probability for each posture. We also use posture-specific user models to predict the probability of the legitimate user per posture. We then combine these probabilities, as shown in Figure 3.

In contrast to a hard decision for a specific posture, our approach always includes all training data from all postures for its predictions, and can thus better respect user-specific characteristics possibly present across postures.

Formally, we estimate the probability $p(h|i)$ of hand posture h given input i (the typing feature vector), and the probability $p(u|h,i)$ of the legitimate user u given posture h and input i. We can then integrate out the posture:

$$p(u|i) = \sum_{h \in H} p(u|h,i)p(h|i), \quad \text{for the set of postures } H$$

Any suitable models can be used to estimate $p(u|h,i)$ and $p(h|i)$. In this paper, we implement this framework as follows: For $p(u|h,i)$, we use the LSAD model, training one such model for each posture. For $p(h|i)$, we train a probabilistic SVM on the owner's data from all three postures.

Type	6 characters	8 characters
dictionary word	monkey	password
pronounceable	Igur39	Bedufo20
random	12hsVi	s5mqde3A

Table 1. Passwords used as stimuli in the user study.

STUDY AND DATA COLLECTION

We collected typing data in a user study with two sessions one week apart. The independent variables were *hand posture* and *password*. As dependent variables, we recorded keystrokes with timestamps and touch locations to derive the described features. Participants also filled in a short questionnaire.

We examined three common postures, all in portrait orientation: 1) THUMB, holding the device in the right hand, touching with the right thumb; 2) TWO-THUMBS, holding it in both hands, touching with both thumbs; and 3) INDEX finger, holding it in the left hand, touching with the right index finger.

We asked participants to repeatedly enter six passwords, shown in Table 1. These passwords were selected to cover two different lengths and three styles: dictionary words, pronounceable passwords [37], and random ones.

Participants

We recruited 28 participants with an average age of 25 years (range: 20-33). 8 were female, 20 male. All were undergraduate or graduate students. One was left-handed. All stated that they own mobile phones with touchscreens and that they (also) employ their right hand for typing. They were compensated with a €15 gift card for an online shop.

Apparatus

We used a Nexus 5 phone. Our app showed the password at the top and the entered text in a box in the centre. We used a custom touchscreen keyboard to measure all features. Style, size and functionality mirrored the default keyboard on the Nexus 5. Thus, capitalisation turned off after one keystroke and vibration issued haptic feedback on touch down.

Procedure

Each participant was invited to two sessions, with a gap of at least one week. Each session comprised three main tasks (three hand postures). The order of tasks varied between subjects according to a 3×3 latin-square design to minimise possible learning effects and fatigue. A different latin-square was used for each session to vary the order of tasks between weeks as well. Sessions lasted for about one hour. Users sat on a couch and were reminded to take breaks.

Between tasks, the app instructed users to assume the correct hand posture for the next task. The phone was held in portrait orientation. Participants were informed to neither put special emphasis on typing speed nor accuracy, but rather to touch naturally as they would do in their usual everyday typing.

For each hand posture, participants typed 6 different passwords in random order, 20 times each. Words were submitted with the return key. To advance to the next repetition, the current one had to be entered correctly without extra key presses. Otherwise, the text was cleared and the user could try again. The number of attempts was unlimited. In total, we collected 2 *sessions* × 28 *users* × 3 *postures* × 6 *passwords* × 20 *repetitions* = 20.160 correct passwords with 201.600 touches.

Feature	Authentication Equal Error Rate (%)								
	THUMB			TWO-THUMBS			INDEX		
	GM	kNN	LSAD	GM	kNN	LSAD	GM	kNN	LSAD
hold time	**31.98**	**32.02**	**30.87**	**26.54**	**26.47**	**25.57**	40.34	40.73	39.24
flight time	35.91	34.52	34.55	32.58	31.64	31.55	**36.92**	**36.85**	**36.60**
up-up	33.95	**32.88**	32.67	**29.91**	**28.77**	**29.45**	**37.13**	**36.71**	**36.99**
down-down	34.44	33.09	33.13	31.62	**30.63**	30.42	**37.31**	**37.04**	37.33
offset x	**33.66**	**33.06**	**31.56**	**31.65**	**31.22**	**29.82**	**39.30**	**38.71**	**37.56**
offset y	**33.28**	**32.66**	**31.31**	**30.45**	**30.07**	**29.45**	**36.21**	**35.79**	**34.07**
down x	34.12	33.55	32.23	32.40	31.42	**30.38**	39.48	39.19	37.63
down y	34.02	33.41	32.27	31.76	31.25	31.08	**36.63**	**36.29**	**35.04**
up x	**33.64**	**33.12**	**31.59**	**31.65**	**31.22**	**29.82**	**39.30**	**38.71**	**37.56**
up y	**33.62**	**33.09**	**31.99**	**31.37**	**30.90**	**31.03**	**36.40**	**36.19**	**34.82**
jump x	35.58	34.80	33.84	32.93	32.29	31.08	38.84	38.43	37.13
jump y	36.92	36.37	34.87	35.49	34.65	35.09	40.02	38.89	38.75
jump angle	37.81	37.76	36.81	33.15	32.52	31.93	39.87	39.41	38.89
jump distance	34.76	34.23	32.17	32.39	32.11	31.65	37.70	37.47	**35.64**
drag x	44.69	45.02	43.45	45.51	45.53	44.22	48.32	48.92	46.43
drag y	45.08	45.56	44.33	45.60	46.00	44.93	46.53	47.24	46.03
drag angle	45.02	45.05	45.05	44.27	44.36	44.32	45.55	45.47	45.41
drag distance	44.09	44.76	44.13	44.14	44.84	42.06	46.21	46.65	44.89
down size	**32.63**	**32.49**	**29.82**	**31.39**	**31.24**	**29.38**	**37.81**	37.51	37.21
up size	34.98	34.76	32.50	33.34	33.23	31.61	40.41	39.94	37.95
down major*	**32.63**	**32.76**	**30.62**	**31.39**	31.67	30.52	**37.81**	**36.17**	**36.03**
up major*	34.98	34.99	32.59	33.34	33.67	32.18	40.41	39.81	38.47
down pressure	**31.38**	**31.03**	**30.32**	**28.59**	**28.91**	**27.84**	**33.14**	**33.32**	**32.61**
up pressure	40.19	39.90	36.37	39.07	39.32	36.05	42.55	42.86	40.37

* The study phone estimated a spherical touch area and therefore returned identical values for major and minor axes.

Table 2. Single feature evaluation. The table shows EERs when using each feature on its own. Highlighted are the top third features (and their x/y counterparts) per model/posture combination. Overall, the best features are hold time, touch down pressure and size, and touch offsets/locations. These results show the potential of touch features.

RESULTS

Since users typed passwords they likely never used before, we considered the first three repetitions of each password as training and removed them from the evaluation. We also removed interrupted entries - those in which at least one flight time exceeded five standard deviations of all repetitions for this user and password (1.6% of each user's data per session).

In general, our evaluation setup assumes each user to be the device "owner" once. A model is trained for this user, either on data from the first session (for evaluation across sessions), or with leave-one-out cross-validation (for within session evaluation). The model is then fed the owner's testing data (i.e. data from second session, or test case from cross-validation) to record predicted probabilities for this legitimate user. Complementary, the model is fed data from all other users to record predicted probabilities for "attackers". After all users have been processed in this way, a threshold is applied to all recorded probabilities to compute global true/false acceptance rates and true/false rejection rates.

We report equal error rates (EER), obtained by varying the threshold applied to the predictions until false acceptance rate and false rejection rate are equal. We chose the EER statistic, since it provides a one-number-summary and is almost always reported in related work. We visualise results with Receiver-Operating-Characteristic (ROC) curves, which plot false positive rates against true positive rates for varying thresholds. Finally, we report comparisons of EERs in *relative* difference (e.g. 4% to 6% EER yields a 50% increase).

Feature Evaluation

To evaluate which touch and typing features are most useful to describe individual behaviour, we first examine the power of single features, before optimising feature sets.

Single Features

To assess each feature's discriminative power alone, we trained different user models on each feature on its own, as shown in Table 2. Across all three tested user models, we observed lowest EERs for hold time, touch down pressure and size, and touch offsets/locations. These results indicate that spatial touch features should be considered for mobile keystroke biometrics. Touch offsets, pressure and size almost always outperformed the flight time commonly used in related work. Typing with both thumbs was more individual than with one, as indicated by lower EERs, while the index finger was the least individual posture. Comparing models, LSAD performed best for almost all feature/posture cases.

Feature Sets

We also evaluated different sets of features to assess their combined potential for user authentication (Table 3). These sets were found with a wrapper approach [35]. Wrappers optimise feature sets by greedily adding the feature which leads to the highest improvement. We applied the wrapper to the best model from single feature evaluation (LSAD), hence the presented results for the other two models can be considered pessimistic. Table 3 shows that our proposed spatial touch features are superior to the temporal features, with 14.3 - 23.5% lower EERs. The best found sets show that offsets were the most useful features based on touch locations.

Combining spatial and temporal features outperformed feature sets consisting of only one of these dimensions: The best combined feature sets achieved 8.5 - 26.3% lower EERs than the best spatial feature sets, and 26.4 - 36.8% lower EERs than the best temporal sets. Thus, spatial and temporal features should be combined for mobile keystroke biometrics.

Although down pressure achieved the lowest EERs for single features (Table 2), it never appeared in the best sets selected from all features. An additional analysis showed that offsets resulted in 5.9 - 12.5% lower EERs than exchanging them for down pressure in the best found sets. This indicates that sensing pressure can be suitably replaced by measuring touch-to-key offsets in this context. Henceforth, we employ the best found feature set for each posture.

Comparison to Related Work

In general, a direct comparison of EERs with related work is difficult due to different devices, tasks and evaluations. Nonetheless, we can compare feature sets used in related work with the features proposed in this paper when evaluated on our data and with our models, see Figure 4. Note that our feature sets were optimised on our data and models. Nevertheless, the comparison shows the benefits of 1) considering touch-specific features and 2) optimising feature sets including touch-specific features - in contrast to related work, where these features were not considered [15, 33, 42, 55, 57] or not tested for optimal combinations [20].

Best Feature Set	**Authentication EER (%)** GM	kNN	LSAD
Spatial			
THUMB: offset x/y, up/down size, jump x	27.38	25.38	20.06
TWO-THUMBS: offset x/y, up/down size	23.35	21.73	18.65
INDEX: offset x/y, up/down size, jump distance	32.27	31.19	26.76
Temporal			
THUMB: hold time, up-up time	28.59	27.75	26.22
TWO-THUMBS: hold time, down-down time	24.40	23.64	21.75
INDEX: hold time, up-up time, flight time	34.57	33.72	33.25
Spatio-Temporal & Pressure			
THUMB: hold time, offset x/y, up/down size	24.32	22.63	17.00
TWO-THUMBS: hold time, offset x/y, up/down size	19.01	17.60	13.74
INDEX: hold time, offset x/y, up/down size, up-up time, jump x, jump distance	30.84	29.48	24.48

Table 3. Feature set evaluation across sessions. The table shows best found feature sets when considering only spatial, only temporal, or all features. These results show that mobile keystroke biometrics benefit from the proposed spatial touch features, including touch-to-key offsets.

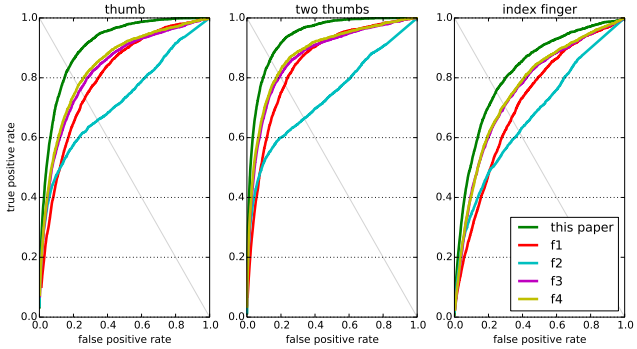

Figure 4. Comparison with feature sets from related work. The figure shows ROC curves obtained when using the LSAD model with the proposed feature sets, and when using it with feature sets occurring in related research (f1 [10, 15, 42], f2 [20], f3 [46], f4 [55, 59]). Crossings with the falling grey lines mark EERs. The results show that touch features should be included when optimising feature sets.

Within Sessions vs Across Sessions

So far, we have conducted evaluations across sessions, since user behaviour can be expected to vary over time. Table 4 now compares analyses within and across sessions: Averaged per posture, anomaly detectors achieved 40.9 - 53.3% lower EERs within sessions than across both sessions. Classifiers achieved 61.6 - 72.8% lower EERs (incl. attacker) and 51.2 - 64.0% lower EERs (excl. attacker) within sessions than across. These results show that evaluation with data collected in a single session is highly optimistic and should thus be avoided to assess the system's expected accuracy in practice.

Classification vs Anomaly Detection Methods

Table 4 also compares the three described training schemes. Averaged over all models per posture for the within session case, classifiers achieved 28.4 - 48.1% lower EERs (excl. attacker) and 45.2 - 58.2% lower EERs (incl. attacker) than anomaly detectors. Complementary, across sessions, classifiers led to 11.7 - 31.1% lower EERs (excl. attacker) and 16.8 - 38.1% lower EERs (incl. attacker). We conclude that classification evaluations yield highly optimistic results if classifiers are not applicable to the intended use-case.

Training Scheme	Equal Error Rate (%)					
	THUMB		TWO-THUMBS		INDEX	
	within	across	within	across	within	across
a) with others (incl. attacker)						
NB	9.71	19.35	10.69	22.06	14.72	37.88
kNNC	3.05	11.15	1.19	8.44	4.03	18.30
SVM	3.43	9.09	1.52	6.71	2.98	14.39
b) with others (excl. attacker)						
NB	9.96	19.58	10.71	22.14	14.78	37.88
kNNC	6.27	13.61	3.47	11.29	7.77	20.09
SVM	5.38	10.87	3.31	8.57	4.85	16.09
c) owner data only						
GM	17.04	24.32	11.71	19.01	19.77	30.84
kNN	14.02	22.63	8.89	17.60	16.55	29.48
LSAD	7.69	17.00	3.84	13.74	10.74	24.48

Table 4. EERs within and across sessions for different training schemes and models. These results reveal two insights: First, due to variance in behaviour over time, error rates obtained within a single session do not appropriately reflect the practically realistic case across sessions. Second, assuming the availability of data from others for training classifiers (cases a and b) leads to highly optimistic results compared to using only data from the legitimate user (c), especially if the specific participant used as an attacker is also included during training (compare a to b).

Comparing both classification variants, including data from the specific attacker led to 20.7 - 25.1% lower EERs (within sessions) and to 4.7 - 11.4% lower EERs (across sessions), averaged over the three tested models per posture. Hence, if classifiers are applicable in a studied use-case, evaluations should still exclude attacker data from the training sets, since it reduces errors but is most likely not available in practice.

Fixed vs Changing Hand Postures

So far, we have separated the data by hand posture, effectively assuming fixed and known postures. However, changing postures have to be considered for usable applications.

To evaluate how password typing behaviour varies between postures, we trained models on data from one posture and tested them with data from a different one. Table 5 shows the resulting EERs for all posture combinations. Averaged over all combinations and models, EERs increased by 86.3%, when models were tested with a different posture. Hence, assuming a fixed posture yields highly optimistic evaluations: Demanding a specific posture (ideally two-thumbs) improves security, but is an undesirable restriction regarding usability.

We have described a probabilistic method to enable more usable keystroke-based authentication with changing hand postures. To evaluate it, we trained models on the data from all three postures from the first session (given the ground-truth postures), and tested them with data from all three postures from the second session (without providing ground-truth postures). This resembles an application asking for all postures during enrolment, but then leaving users free to type with any of the enrolled postures without telling the system which one they are currently using.

Our proposed probabilistic framework, implemented as described in this paper, achieved 21.02% EER. Note that here we test with data from all postures. Therefore, this value also fits the expectation that a system's extension, which suitably handles changing postures, should lead to EERs between those obtained for fixed single postures (LSAD: two-thumbs 13.74% to index 24.48%). In the following, we compare our system to two simple alternatives.

Authentication Equal Error Rate (%) Across Hand Postures			
	THUMB	TWO-THUMBS	INDEX
GM			
THUMB	24.32	38.67	40.41
TWO-THUMBS	35.66	19.01	40.76
INDEX	43.29	44.13	30.84
kNN			
THUMB	22.63	38.29	39.82
TWO-THUMBS	34.85	17.60	40.33
INDEX	42.88	43.88	29.48
LSAD			
THUMB	17.00	35.48	37.24
TWO-THUMBS	33.05	13.74	39.12
INDEX	42.45	59.07	24.48

Table 5. Equal error rates across sessions when using data of different hand postures for training (rows) and testing (columns). These results show that mobile keystroke-based biometrics are highly posture-specific.

Across postures, our framework yields a reduction in EERs by 36.4 - 64.4% compared to the values in Table 5, which correspond to a system naïvely assuming that training data from any posture is suitable for predictions for all postures.

We also compared our method to models trained on data from all postures pooled together: The best EER obtained was 27.38%, using the kNN model. In comparison, our proposed method thus reduced EER by 23.2%.

In conclusion, our probabilistic framework handles changing hand postures better than approaches, which either 1) simply ignore differences between postures (Table 5), or 2) just train a single model. In contrast, our framework estimates the probability of the legitimate user *for each posture*, in addition to the probability of each posture being indeed the one used.

SUMMARY AND IMPLICATIONS

In this section, we summarise challenges and opportunities and present implications derived from our analyses.

Challenges: We quantified the following challenges for practical and usable applications of mobile keystroke dynamics:

- *Mobile typing biometrics vary over time:* Training and testing models on data collected in a single session results in EERs less than half of the values observed across sessions.
- *Data from multiple users improves authentication accuracy, but is not applicable to password-hardening:* Evaluation with classifiers led to 11.7 - 31.1% lower EERs than employing anomaly detectors. However, classifiers using feature-vectors from different users typing a specific password impractically imply that the word is known to others.
- *Mobile typing biometrics are highly hand posture-specific:* Training and testing models on different postures increased EERs by 86.3% relative to testing with the same posture.

These challenges imply three important considerations for applicable and usable mobile keystroke dynamics: First, user studies should always include multiple sessions for each participant. Second, classifiers should only be used in evaluations if they are also applicable to the targeted threat model. This is not the case if passwords have to be revealed to others. Finally, applications of mobile keystroke biometrics have to infer postures dynamically to retain usability.

Opportunities: To improve keystroke biometrics for usable applications, we revealed and quantified these opportunities:

- *Spatial touch features outperform the traditional temporal features:* The best spatial feature sets reduced EERs by 14.3 - 23.5% relative to the best temporal sets.
- *Spatial touch features outperform pressure features:* Pressure never appeared in optimised feature sets, while offsets were always selected. They resulted in 5.9 - 12.5% lower EERs than exchanging them for pressure in the best sets.
- *Spatial and temporal features complement each other:* The best spatio-temporal feature sets reduced EERs by 8.5 - 26.3% relative to the best spatial sets, and by 26.4 - 36.8% relative to the best temporal sets.
- *Models for different hand postures can be combined to allow for changing postures:* Combining models in a probabilistic framework reduced EERs by 23.2% compared to training a single model on data from all postures, and by 36.4 - 64.4% compared to ignoring posture differences.

In consequence, we recommend to measure touch-to-key offsets to improve accuracy of mobile keystroke biometrics. Furthermore, these features should be combined with the temporal typing features known from related work. Finally, hand postures present a trade-off between security and usability: A fixed posture results in lower EERs, but restricts the user. This can be addressed by combining posture-specific models.

We expect lessons learned in this work to be useful on a broader scale beyond keystroke biometrics: Variability over time and between postures can be expected for other mobile biometrics (e.g. general touch behaviour). Moreover, potential problems regarding the use of classifiers should be considered for all behavioural biometrics related to secret tokens, not just passwords (e.g. PINs, shapes, gestures).

DISCUSSION

On our collected study data, we reduced EERs by up to 36.8% with the proposed feature sets, including touch-specific features. While a direct comparison of absolute EERs with related work is difficult due to different devices, tasks and evaluations, we showed improvements compared to feature sets employed in related work when tested with our models and data. To further improve absolute accuracy, practical systems could combine touch features with others (e.g. motion [59]), or observe mobile device usage more holistically, with typing biometrics as one part of it (multimodal biometrics [17]).

For privacy reasons, the data processing system should run on the device, not in a cloud. We analysed the study data on a PC, but also measured crucial operations on the device: Training an LSAD model is determined by computing and inverting a kernel matrix [44]. For our training set sizes and number of features, the required matrix operations took 50 ms on the Nexus 5. Authentications after password entry are simple vector multiplications, unnoticeable to users.

We highlighted that classifiers are not applicable to keystroke biometrics for password entry, if features are extracted for the specific word used: Since classifiers are trained on data from multiple users, their use implies that the password was entered by others and is thus not secret any more.

However, we do not argue against the use of classifiers in general: In this paper, we only discussed a *static* task, meaning that the system expects a fixed text (i.e. password) to be entered. For *dynamic* tasks (i.e. free text entry), typing behaviour can be described for common words [11] or for bigrams [9, 50]. In these cases, it is reasonable to assume recorded data from others to train classifiers. While we did not explicitly cover such a dynamic task in this paper, our results with classifiers show that they also benefit from the proposed touch-specific typing features.

Our analysis revealed that typing behaviour is highly posture-specific. This presents a challenge to usable applications of mobile keystroke biometrics. We proposed a probabilistic framework to allow users to type with changing postures. It outperformed models trained across all postures as well as "posture-agnostic" approaches.

LIMITATIONS

We only collected right-handed touches, limiting the observed set of postures. Our evaluations across postures can assess accuracy for changing postures between entries, or between enrolment and entry. However, they can neither assess posture changes mid-typing nor continuous changes, which might occur, for example, due to hand drift [13].

A set of three anomaly detectors and three classifiers was evaluated with six different passwords. Many more methods exist [50], and could be tested with a broader set of passwords to improve generalisability. Nevertheless, our results suggest that the proposed typing features can be suitably used by different models.

We conducted a lab study with participants sitting down in two sessions. An "in the wild" study may observe greater variability in long-term behaviour with varying contexts and phone models. Regarding touch features, related work showed that offset patterns are to some extent robust across changing conditions, and that they are highly individual on other phone models as well [12, 54].

CONCLUSION AND FUTURE WORK

In this paper, we have studied mobile-specific and touch-specific challenges and opportunities for keystroke biometrics. We have revealed, analysed and discussed different improvements for a password entry use-case and threat model. These analyses and improvements are important to advance keystroke biometrics on mobile touchscreen devices.

Overall, our results 1) improve implicit authentication accuracy through new features, 2) support realistic evaluations leading to applicable systems, and 3) improve usability with a framework to handle changing hand postures.

In summary, we first complemented existing temporal typing features with touch-specific spatial features. Second, we quantified the effect of three common assumptions, namely within-session evaluation, training on attacker data, and assuming fixed hand postures. We revealed that these practices can result in overly optimistic assessments with respect to usable applications in practice. In consequence, we addressed these issues with evaluation recommendations and a probabilistic method to account for changing hand postures.

We plan to use touch-specific features in a dynamic typing task, such as free text messaging. While related work [9, 50] extracted features per bigram in such cases (e.g. flight time from *t* to *h*), in other words features for discrete key-combinations, we will instead train regression models to map precise touch locations to feature values [12, 54].

PROJECT RESOURCES

Please contact the first author to gain access to the dataset.

ACKNOWLEDGEMENTS

This work was partially supported by funding from a Google Research Award.

REFERENCES

1. Angulo, J., and Wästlund, E. Exploring Touch-Screen Biometrics for User Identification on Smart Phones. *Privacy and Identity Management for Life 375* (2012), 130–143.
2. Aviv, A., Gibson, K., Mossop, E., Blaze, M., and Smith, J. Smudge Attacks on Smartphone Touch Screens. In *WOOT* (2010), 1–7.
3. Azenkot, S., and Zhai, S. Touch Behavior with Different Postures on Soft Smartphone Keyboards. In *MobileHCI 2012* (2012), 251–260.
4. Baldwin, T., and Chai, J. Towards Online Adaptation and Personalization of Key-Target Resizing for Mobile Devices. In *IUI 2012* (2012), 11–20.
5. Banerjee, S., and Woodard, D. Biometric Authentication and Identification using Keystroke Dynamics: A Survey. *Journal of Pattern Recognition Research 7* (2012), 116–139.
6. Baudisch, P., and Chu, G. Back-of-Device Interaction Allows Creating Very Small Touch Devices. In *CHI 2009* (2009), 1923–1932.
7. Beton, M., Marie, V., and Rosenberger, C. Biometric Secret Path for Mobile User Authentication: A Preliminary Study. In *WCCIT 2013*, Ieee (June 2013), 1–6.
8. Böhmer, M., Hecht, B., Schöning, J., Krüger, A., and Bauer, G. Falling Asleep with Angry Birds, Facebook and Kindle - A Large Scale Study on Mobile Application Usage. *MobileHCI 2011* (2011), 47–56.
9. Bours, P., and Barghouthi, H. Continuous Authentication Using Biometric Keystroke Dynamics. In *NISK* (2009), 1–12.
10. Buchoux, A., and Clarke, N. L. Deployment of Keystroke Analysis on a Smartphone. In *Australian Information Security Management Conference* (2008).
11. Burgbacher, U., and Hinrichs, K. An Implicit Author Verification System for Text Messages Based on Gesture Typing Biometrics. In *CHI 2014* (2014), 2951–2954.
12. Buschek, D., Rogers, S., and Murray-Smith, R. User-Specific Touch Models in a Cross-Device Context. In *MobileHCI 2013* (2013), 382–391.
13. Buschek, D., Schoenleben, O., and Oulasvirta, A. Improving Accuracy in Back-of-Device Multitouch Typing: A Clustering-based Approach to Keyboard Updating. In *IUI 2014* (2014), 57–66.
14. Campisi, P., Maiorana, E., Lo Bosco, M., and Neri, A. User authentication using keystroke dynamics for cellular phones. *IET Signal Processing 3*, 4 (2009), 333–341.
15. Clarke, N. L., and Furnell, S. M. Authenticating mobile phone users using keystroke analysis. *International Journal of Information Security 6*, 1 (Aug. 2006), 1–14.
16. Crawford, H. Keystroke Dynamics: Characteristics and Opportunities. In *8th International Conference on Privacy, Security and Trust* (2010), 205–212.
17. Crawford, H. A. *A Framework for Continuous, Transparent Authentication on Mobile Devices*. PhD thesis, University of Glasgow, 2012.
18. De Luca, A., Hang, A., Brudy, F., Lindner, C., and Hussmann, H. Touch me once and I know its you! Implicit Authentication based on Touch Screen Patterns. In *CHI 2012* (2012), 987–996.
19. Devroye, L., Györfi, L., and Lugosi, G. *A Probabilistic Theory of Pattern Recognition*. Springer, 1996.
20. Draffin, B., Zhu, J., and Zhang, J. KeySens: Passive User Authentication through Micro-behavior Modeling of Soft Keyboard Interaction. *Mobile Computing, Applications, and Services 130* (2014), 184–201.
21. Edelmann, J., Mock, P., Schilling, A., Gerjets, P., Rosenstiel, W., and Strasser, W. Towards the Keyboard of Oz: Learning Individual Soft-Keyboard Models from Raw Optical Sensor Data. In *ITS 2012* (2012), 163–172.
22. Forlines, C., Wigdor, D., Shen, C., and Balakrishnan, R. Direct-Touch vs. Mouse Input for Tabletop Displays. In *CHI 2007* (2007), 647–656.
23. Frank, M., Biedert, R., Ma, E., Martinovic, I., and Song, D. Touchalytics: On the Applicability of Touchscreen Input as a Behavioral Biometric for Continuous Authentication. *Information Forensics and Security 8*, 1 (2013), 126–148.
24. Goodman, J. T., Venolia, G., Steury, K., and Parker, C. Language Modeling for Soft Keyboards. In *IUI 2002* (Jan. 2002), 194–195.
25. Harbach, M., von Zezschwitz, E., Fichtner, A., De Luca, A., and Smith, M. It's a Hard Lock Life: A Field Study of Smartphone (Un) Locking Behavior and Risk Perception. In *SOUPS 2014* (2014), 213–230.
26. Henze, N., Rukzio, E., and Boll, S. 100,000,000 Taps: Analysis and Improvement of Touch Performance in the Large. In *MobileHCI 2011* (2011), 133–142.
27. Holz, C., and Baudisch, P. The Generalized Perceived Input Point Model and How to Double Touch Accuracy by Extracting Fingerprints. In *CHI 2010* (2010), 581–590.

28. Holz, C., and Baudisch, P. Understanding Touch. In *CHI 2011* (2011), 2501–2510.

29. Holz, C., and Baudisch, P. Fiberio: A Touchscreen that Senses Fingerprints. In *UIST 2013* (2013), 41–50.

30. Hwang, S.-s., Cho, S., and Park, S. Keystroke dynamics-based authentication for mobile devices. *Computers & Security 28*, 1-2 (Feb. 2009), 85–93.

31. Jakobsson, M., Shi, E., Golle, P., and Chow, R. Implicit Authentication for Mobile Devices. In *HotSec 2009* (2009).

32. Jeanjaitrong, N., and Bhattarakosol, P. Feasibility Study on Authentication Based Keystroke Dynamic over Touch-Screen Devices. In *ISCIT 2013*, Ieee (Sept. 2013), 238–242.

33. Karatzouni, S., and Clarke, N. Keystroke Analysis for Thumb-based Keyboards on Mobile Devices. In *IFIP International Federation for Information Processing*, vol. 232 (2007), 253–263.

34. Killourhy, K. S., and Maxion, R. A. Comparing Anomaly-Detection Algorithms for Keystroke Dynamics. In *Dependable Systems & Networks* (June 2009), 125–134.

35. Kohavi, Ron and John, G. H. Wrappers for feature subset selection. *Artificial Intelligence 97*, 1 (1997), 273–324.

36. Kristensson, P. O. *Discrete and Continuous Shape Writing for Text Entry and Control*. PhD thesis, Linköping University, Sweden, 2007.

37. Lau, S.-h., Siena, S., Pandey, A., Sosothikul, S., Cranor, L., Ur, B., and Shay, R. Exploring the Usability of Pronounceable Passwords. In *SOUPS Poster* (2014).

38. Maiorana, E., Campisi, P., González-Carballo, N., and Neri, A. Keystroke Dynamics Authentication for Mobile Phones. In *SAC 2011* (2011), 21–26.

39. Mock, P., Edelmann, J., Schilling, A., and Rosenstiel, W. User Identification Using Raw Sensor Data From Typing on Interactive Displays. In *IUI 2014* (2014), 67–72.

40. Monrose, F., Reiter, M., and Wetzel, S. Password Hardening Based on Keystroke Dynamics. In *Int'l. Journal of Information Security* (2002), 69–83.

41. Monrose, F., and Rubin, A. Authentication via Keystroke Dynamics. In *Computer and Communications Security* (1997), 48–56.

42. Nauman, M., Ali, T., and Rauf, A. Using trusted computing for privacy preserving keystroke-based authentication in smartphones. *Telecommunication Systems 52* (2013), 2149–2161.

43. Platt, J. C. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. *Advances in Large Margin Classifiers 10*, 3 (1999), 61–74.

44. Quinn, J., and Sugiyama, M. A least-squares approach to anomaly detection in static and sequential data. *Pattern Recognition Letters* (2014).

45. Sae-Bae, N., Ahmed, K., Isbister, K., and Memon, N. Biometric-Rich Gestures: A Novel Approach to Authentication on Multi-touch Devices. In *CHI 2012* (2012), 977–986.

46. Saevanee, H., and Bhattarakosol, P. Authenticating user using keystroke dynamics and finger pressure. In *6th IEEE Consumer Communications and Networking Conference* (2009).

47. Saevanee, H., Clarke, N., Furnell, S., and Biscione, V. Text-Based Active Authentication for Mobile Devices. *ICT Systems Security and Privacy Protection 428* (2014), 99–112.

48. Schaub, F., Deyhle, R., and Weber, M. Password Entry Usability and Shoulder Surfing Susceptibility on Different Smartphone Platforms. In *MUM 2012* (2012), 13:1–13:10.

49. Shi, E., Niu, Y., Jakobsson, M., and Chow, R. Implicit Authentication through Learning User Behavior. *Information Security* (2011), 99–113.

50. Teh, P. S., Teoh, A. B. J., and Yue, S. A Survey of Keystroke Dynamics Biometrics. *The Scientific World Journal 2013* (2013).

51. von Zezschwitz, E., Koslow, A., De Luca, A., and Hussmann, H. Making Graphic-Based Authentication Secure against Smudge Attacks. In *IUI 2013* (2013), 277–286.

52. Wang, F., and Ren, X. Empirical Evaluation for Finger Input Properties In Multi-touch Interaction. In *CHI 2009* (2009), 1063–1072.

53. Weir, D., Buschek, D., and Rogers, S. Sparse Selection of Training Data for Touch Correction Systems. In *MobileHCI 2013* (2013), 404–407.

54. Weir, D., Rogers, S., Murray-Smith, R., and Löchtefeld, M. A User-Specific Machine Learning Approach for Improving Touch Accuracy on Mobile Devices. In *UIST 2012* (2012), 465–476.

55. Xu, H., Zhou, Y., and Lyu, M. Towards Continuous and Passive Authentication via Touch Biometrics: An Experimental Study on Smartphones. *SOUPS 2014* (2014), 187–198.

56. Yin, Y., Ouyang, T. Y., Partridge, K., and Zhai, S. Making Touchscreen Keyboards Adaptive to Keys, Hand Postures, and Individuals - A Hierarchical Spatial Backoff Model Approach. In *CHI* (2013), 2775–2784.

57. Zahid, S., Shahzad, M., Khayam, S. A., and Farooq, M. Keystroke-Based User Identification on Smart Phones. In *LNCS*, vol. 5758 (2009), 224–243.

58. Zhao, X., Feng, T., and Shi, W. Continuous Mobile Authentication Using A Novel Graphic Touch Gesture Feature. In *BTAS 2013* (2013), 1–6.

59. Zheng, N., Bai, K., Huang, H., and Wang, H. You Are How You Touch: User Verification on Smartphones via Tapping Behaviors. In *Tech. Rep. WM-CS-2012-06* (2012).