CS 528: Ubiquitous and Mobile Computing

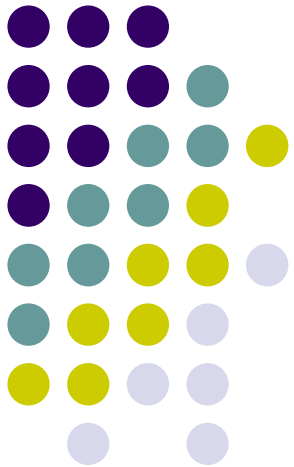
Improving Accuracy, Applicability and Usability of Keystroke Biometrics on Mobile Touchscreen Devices

Kewen Gu

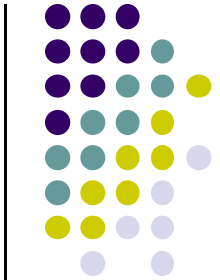
Yuheng Huo

Computer Science Dept.

Worcester Polytechnic Institute (WPI)

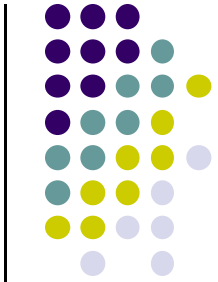


Improving Accuracy, Applicability and Usability of Keystroke Biometrics on Mobile Touchscreen Devices



- Introduction
- Related Work
- Methodology
- Results
- Discussion
- Limitations
- Conclusion and Future Work

Introduction



- Utilizing of keystroke biometrics on mobile touchscreen devices
- For password entry, adding an implicit security layer to observe typing behavior
- Even the password entered is correct, access can still be denied due to different typing behavior

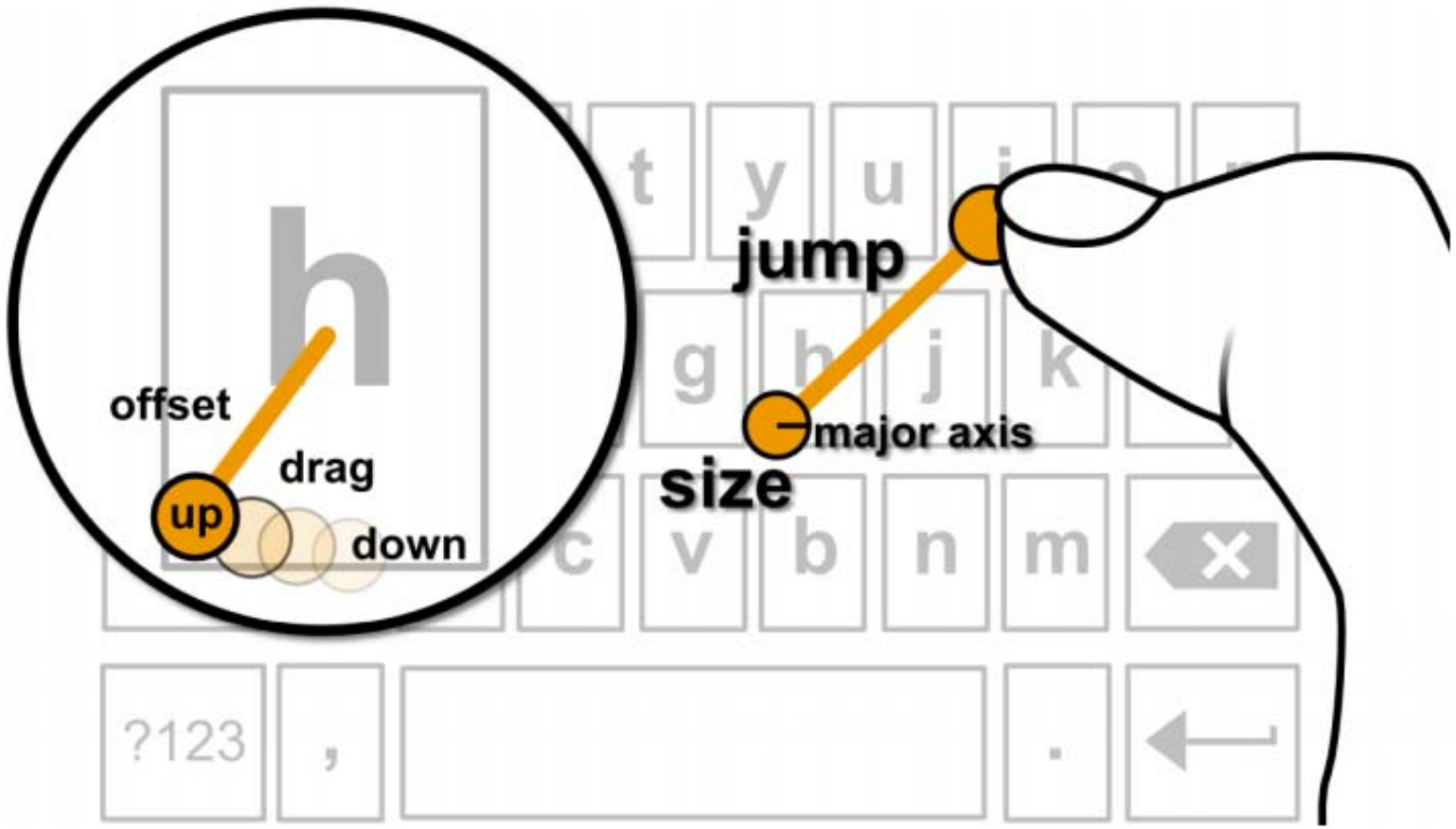
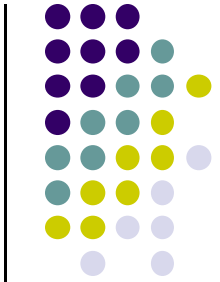


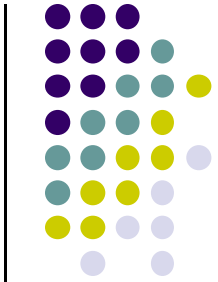
Figure 1. Touch-specific keystroke features on a mobile keyboard. In this example, the user is typing "hi". The magnified *h*-key shows touch down and up locations, the drag in between, and the offset to the key's centre. The keyboard-overlay shows touch area size and axis, as well as the "jump" vector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

Introduction



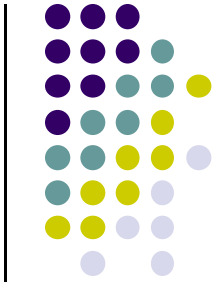
- To improve implicit authentication accuracy
  - Evaluated touch-specific features for capturing individual typing behavior.
  - Spatial touch features outperform the commonly used temporal features
  - Both spatial and temporal features combined to reduce equal error rates by up to 36.8%

Introduction



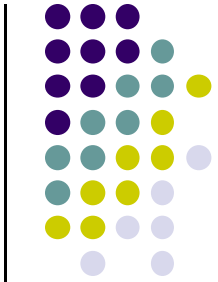
- To improve applicability
  - Discussed and quantified practical implications of different commonly used evaluations
  - Compared result for:
    - Training and testing within sessions or across sessions
    - Training on owner data only or also on data from others
    - Assuming fixed or changing hand postures
  - Allowing for a more realistic assessment of keystroke biometrics in practice

Introduction



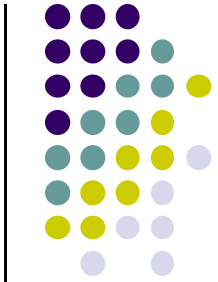
- To improve usability
  - Proposed an approach to avoid restricting users to one typing posture
  - Analyzed one-thumb, two-thumb and index finger typing
  - Showed that behavior is highly posture-specific
  - Presented a method to handle changing hand postures

Related Work



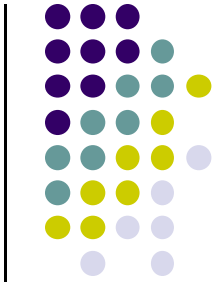
- Modeling Touch and Typing Behavior
  - Related research reduced typing errors with keyboard personalization based on users' individual touch distributions per key
  - Research revealed individual touch typing patterns, but none of these projects utilized this information for user authentication
  - Research also found several influences on general touch behavior

Related Work



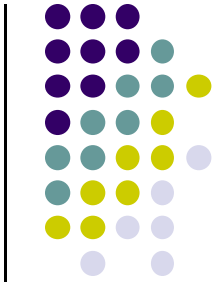
- Behavioral Biometrics for Mobile Typing
  - Related work applied keystroke dynamics on mobile phones with physical keys
  - Used neural networks to authenticate mobile phone users based on temporal typing features
  - Used keystroke latencies and key-hold times, but no touch features

Related Work



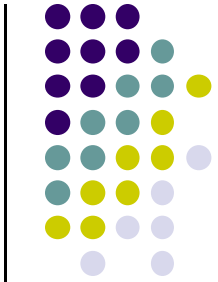
- Touch-based Implicit Authentication and Identification
  - Related work addressed verifying user identity with diverse touch measures
  - Other work suggested to directly replace passwords with touch evidence
  - Further related research distinguished users with rear projected tabletop systems

Related Work



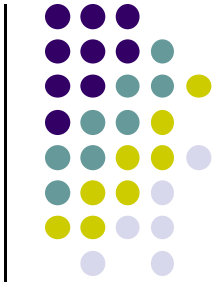
- Opportunities and Intended Contribution
  - Related work on keyboard personalization, targeting, and touch-based authentication has shown individual touch and typing behavior
  - Research on keystroke biometrics has either ignored spatial touch-specific typing features on mobile devices, or only used such features on tabletops or with gesture-keyboards
  - Related work has revealed the need to address mobile applications of keystroke biometrics and to develop novel features

Methodology



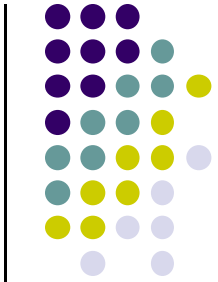
- Thread Model
  - Consider that an attacker gains access to an unlocked device and additionally knows the owner's password
  - Here, keystroke information serves as an additional security layer
  - Even if the attacker enters the correct password, the system can deny access due to different typing behavior

Methodology



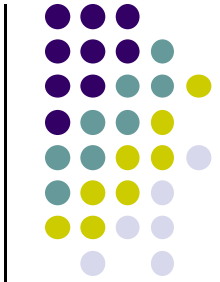
- Improving accuracy of keystroke biometrics
  - Typing features
    - Needed to capture individual aspects of typing behavior in order to build user model for authentication
    - Proposed to consider new spatial touch-specific features:
      - Exact touch locations at touch down and up events
      - Offsets between touch up and key-centers
      - Touch ‘jumps’, the distances between subsequent touches
      - Drag distances/angles between touch down and up locations
      - Touch are sizes, ellipses axes, touch pressure

Methodology



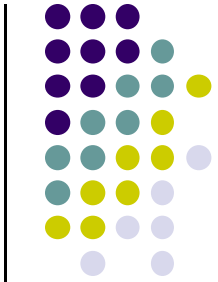
- Improving accuracy of keystroke biometrics
  - User Models for Authentication
    - Compared models of two types:
      - **Anomaly detector**, which only require training data from the legitimate user
      - **Classification methods**, which are trained on data from multiple users
    - In practice, training data can be collected in an enrolment phase or from normal use.
    - During testing, these models decide whether the password was typed by the legitimate user

Methodology



- Improving applicability and usability
  - Evaluation within sessions vs across sessions
    - Training and testing on data from the same session is too optimistic, since, in practical applications, enrolment and authentication will never follow directly one after the other
    - 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

Methodology



- Improving applicability and usability
  - Classification vs anomaly detection methods
    - Classifiers are potentially more powerful, since they characterize the legitimate user in contrast to others, whereas anomaly detectors can only check for deviation from the legitimate user's behavior.
    - Splitting the data for evaluation of classifiers into three parts: owner, attacker, and others (excluding the attacker). Classifiers can be trained on data from owner and others, without assuming known data from the attacking individual.
    - Classifiers lead to 28.4 - 48.1% lower EERs relative to anomaly detectors, and to 45.2 - 58.2% lower EERs

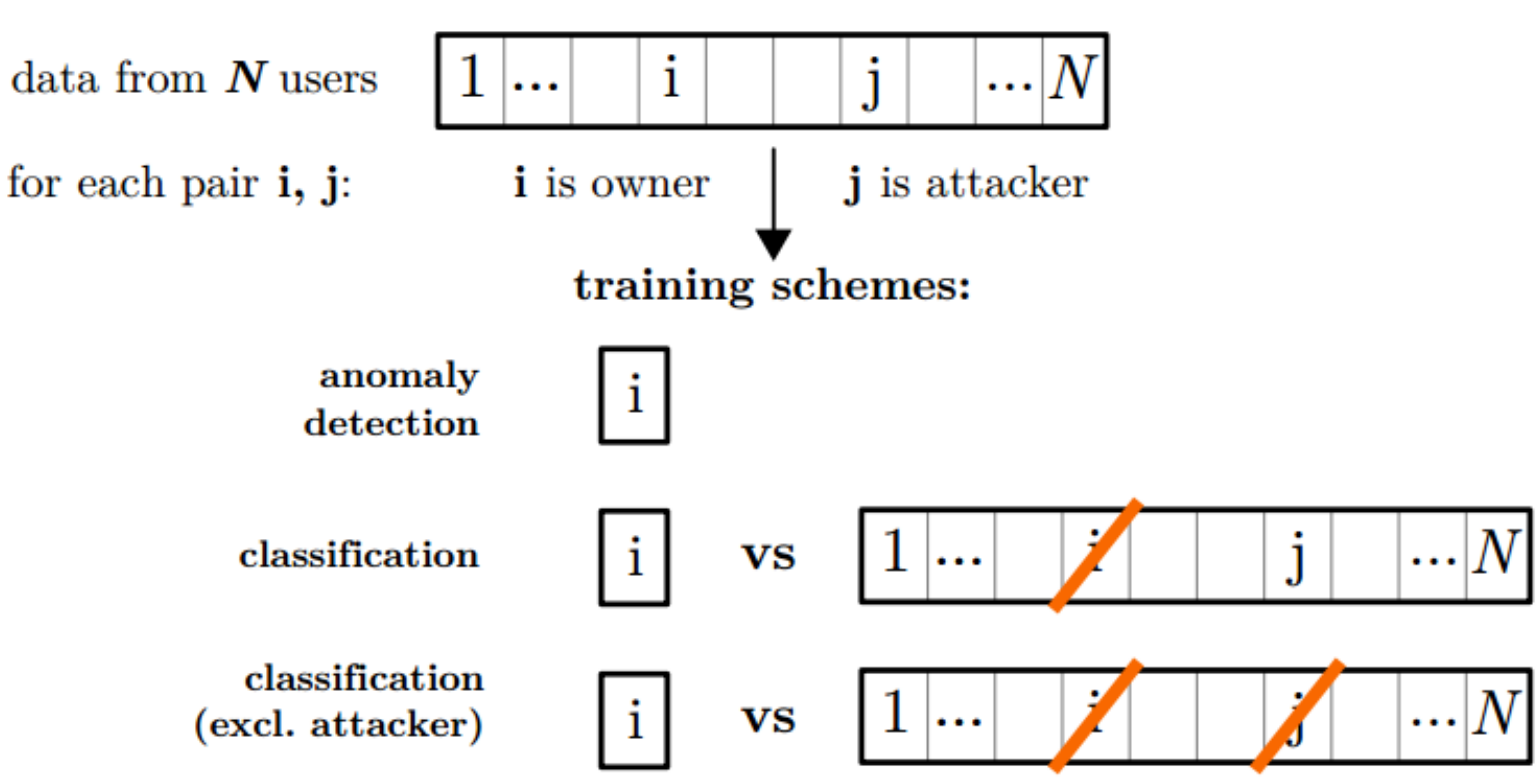
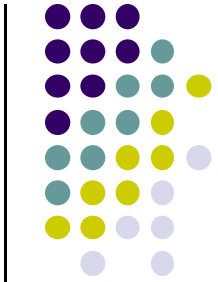


Figure 2. Three *training schemes for evaluation* of keystroke biometrics. Anomaly detectors are only trained on data from the legitimate user ("owner"). In contrast, classifiers also use pooled data from all other users. However, applications may not always have access to typing behaviour of other users 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.

Methodology



- Improving applicability and usability
  - Fixed vs changing hand postures
    - Systems that require fixed hand postures restrict the user's freedom and lowers the usability
    - Proposed a framework that allows users to type with different postures
    - Used a probabilistic classifier to predict a probability for each posture, and use posture-specific user models to predict the probability of the legitimate user per posture, e then combine these probabilities
    - Showed 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.

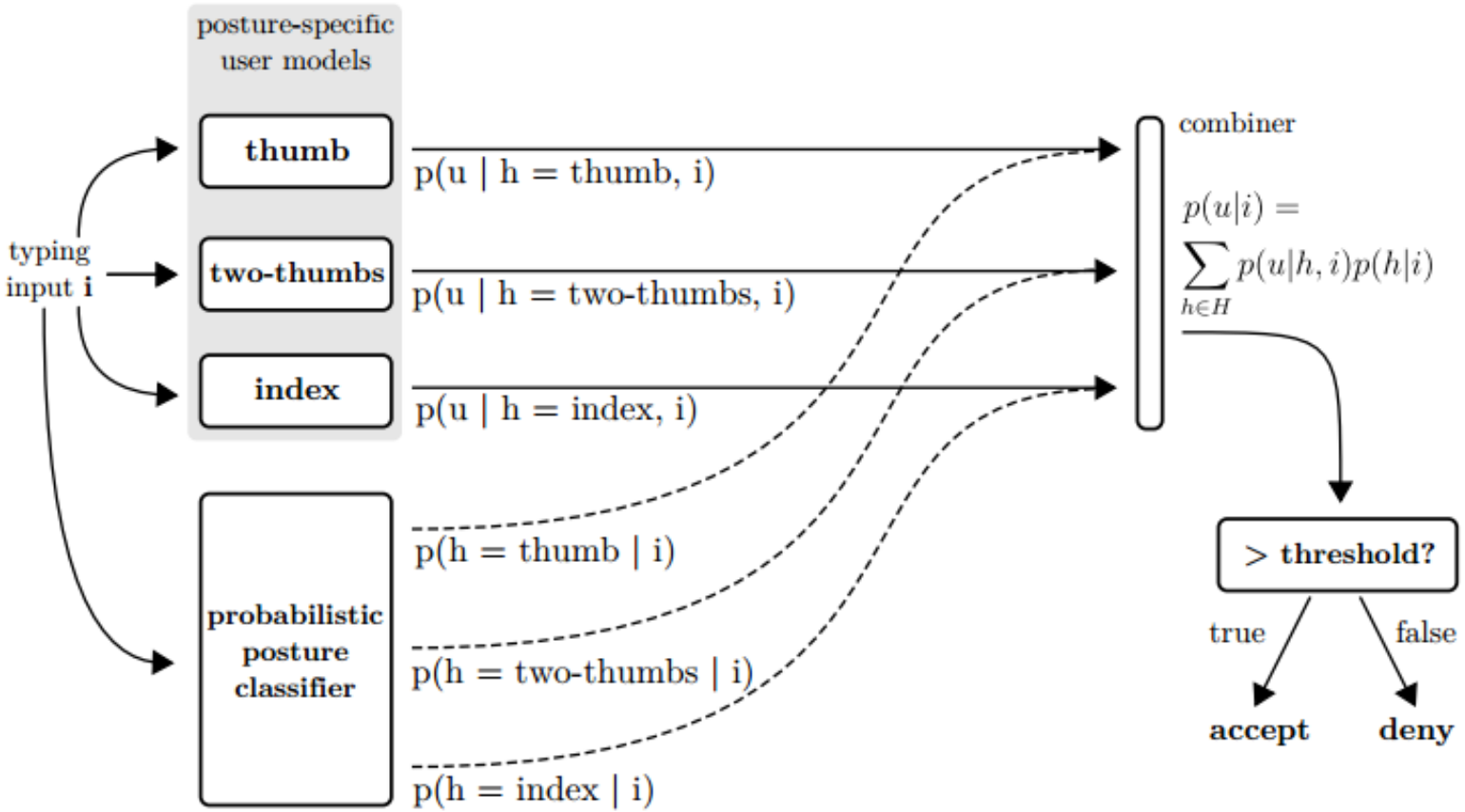


Figure 3. Probabilistic framework for usable mobile keystroke biometrics which does not restrict users to a fixed hand posture. For an entered password (input i), each posture-specific model predicts the probability $p(u|h,i)$ of the legitimate user u, assuming that the corresponding hand posture h was 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.

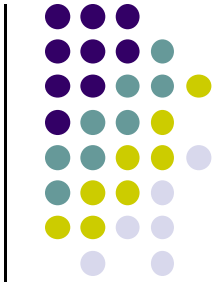
Methodology

- Study and data collection
  - Collected typing data in a user study with two sessions one week apart
  - 28 participants with an average age of 25 years old, in the range of 20 to 33 years old. 8 were female, and 20 male. All were undergraduate or graduate students.
  - All used Nexus 5 phones

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

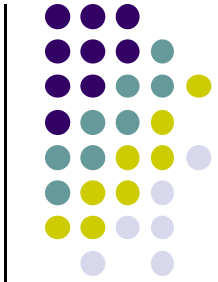
Table 1. Passwords used as stimuli in the user study.

Methodology



- Study and data collection
  - Examined three common postures:
    - Thumb, holding the device in the right hand, touching with the right thumb
    - Two-thumbs, holding it in both hands, touching with both thumbs
    - Index finger, holding it in the left hand, touching with the right index finger
  - For each hand posture, participants typed 6 different passwords in random order, 20 times each. In total, collected 2 sessions × 28 users × 3 postures × 6 passwords × 20 repetitions = 20,160 correct passwords with 201,600 touches

Results



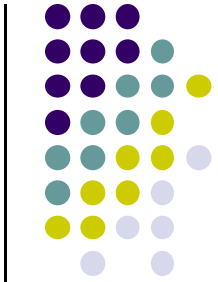
- Feature Evaluation:
  - Two feature evaluations
    - Evaluated which touch and typing feature are most useful to identify individual behavior
  - Single Feature Evaluation:
    - Training models with single feature
    - As shown in table 2, across all three tested user models: hold time, touch pressure, and touch location is observed for lowest EERs. As result, touch feature such as touch pressure, location and size is more unique to individuals than hold time does.

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

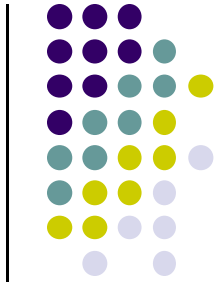
* The study phone estimated a spherical touch area and therefore returned identical values for major and minor axes.

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



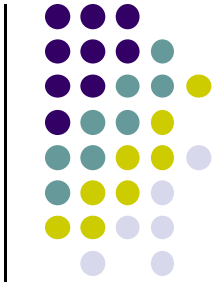
- Feature Set Evaluation:
  - Different sets of features are used to train the models.
    - As shown in table 3, spatial touch features are better than temporal feature, as the authentication EERs of spatial touch features is 14.3 ~ 23.5% lower than temporal features. Combination of both spatial and temporal feature sets achieves 8.5 ~ 16.3% lower EERs than spatial touch features.



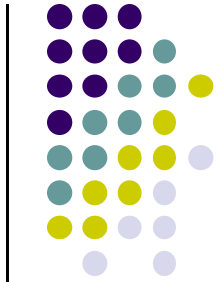
Best Feature Set	Authentication EER (%) GM	kNN	LSAD
Spatial			
THUMB: offset x/y, up/down size, jump x	27.38	25.38	20.06
TWO-THUMBS: offset x/y, up/down size	23.35	21.73	18.65
INDEX: offset x/y, up/down size, jump distance	32.27	31.19	26.76
Temporal			
THUMB: hold time, up-up time	28.59	27.75	26.22
TWO-THUMBS: 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.

Results



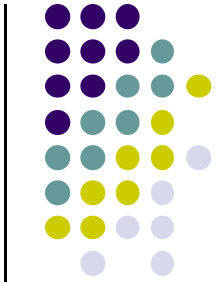
- Fixed vs Changing Hand Postures:
  - Pervious models were trained by assuming known and fixed hand posture. However, dynamic posture should be considered.
  - Table 5 shows when train a model with one posture data and test it with a different posture data, the result average EERs increased by 86.3%. This proves that the model trained is highly posture specific.



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

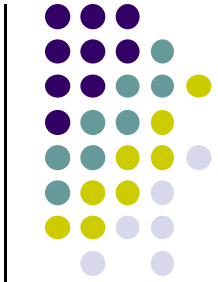
Table 5. Equal error rates across sessions when using data of different hand postures for training (rows) and testing (columns). These results show that mobile keystroke-based biometrics are highly posture-specific.

Results



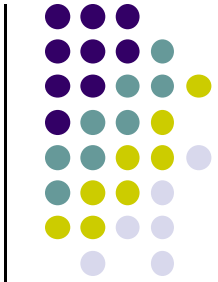
- Fixed vs Changing Hand Postures:
  - The probabilistic framework was implemented to enable changing hand postures.
  - To evaluate it, the framework was trained with all the posture data from the first session, and tested them with data collected from the second session. As result an equal error rate with 21.02% is achieved. This also yielded a reduction in EERs by 36.4 - 64.4% compared to the values in Table 5.

Discussion



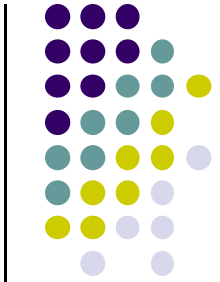
- Challenges:
  - Mobile typing biometrics vary over time
  - Data from multiple users improves authentication accuracy, but is not applicable to password-hardening
  - Mobile typing biometrics are highly hand posture-specific

Discussion



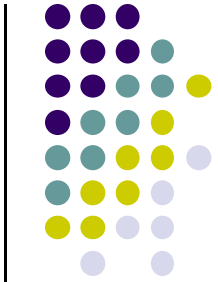
- These challenges imply three important considerations for applicable and usable mobile keystroke dynamics:
  - First, user studies should always include multiple sessions for each participant
  - Second, classifiers should only be used in evaluations if they are also applicable to the targeted threat model.
  - Finally, applications of mobile keystroke biometrics have to infer postures dynamically to retain usability

Discussion



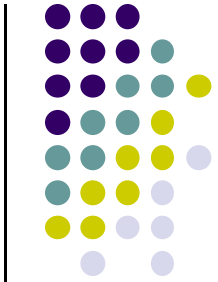
- Opportunities
  - Spatial touch features outperform the traditional temporal features
  - Spatial touch features outperform pressure features
  - Spatial and temporal features complement each other
  - Models for different hand postures can be combined to allow for changing postures

Discussion



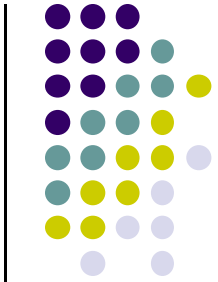
- In this study, EERs were reduced by up to 36.8% with the proposed feature sets.
- It showed improvements compared to feature sets employed in related work when tested with our models and data.
- For privacy reasons, the data processing system run on the device, not in a cloud.

Limitations



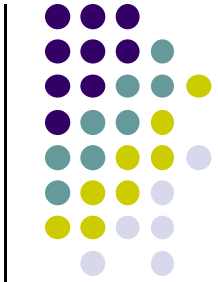
- Only collected right-handed touches, limiting the observed set of postures
- Many more methods exist, and could be tested with a broader set of passwords to improve generalizability
- An “in the wild” study may observe greater variability in long-term behavior with varying contexts and phone models

Conclusion and future work



- This paper revealed, analyzed and discussed different improvements for a password entry use-case and threat model. The results include:
  - Improved implicit authentication accuracy through new features
  - Supported realistic evaluations leading to applicable systems
  - Improved usability by implementing a framework to handle changing hand postures

Conclusion and future work



- Future work:
  - Use touch-specific features in a dynamic typing task, such as free text messaging
  - Train regression models to map precise touch locations with feature values

References

- Azenkot, S., and Zhai, S. Touch Behavior with Different Postures on Soft Smartphone Keyboards. In MobileHCI 2012 (2012), 251–260.
- Crawford, H. A. A Framework for Continuous, Transparent Authentication on Mobile Devices. PhD thesis, University of Glasgow, 2012.
- 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.
- Bohmer, M., Hecht, B., Sch ¨ oning, J., Kr ¨ uger, 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.
- Bours, P., and Barghouthi, H. Continuous Authentication Using Biometric Keystroke Dynamics. In NISK (2009), 1–12.
- 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.
- Jakobsson, M., Shi, E., Golle, P., and Chow, R. Implicit Authentication for Mobile Devices. In HotSec 2009 (2009).