Mobile Biometrics: Towards A Comprehensive Evaluation Methodology

Attaullah Buriro*, Zahid Akhtar‡, Bruno Crispo*, and Sandeep Gupta*
*Department of Information Engineering and Computer Science (DISI),
University of Trento, Via Sommarive, 38123, Italy,
Email*:{attaullah.buriro, bruno.crispo, sandeep.gupta}@unitn.it
‡INRS-EMT, University of Quebec, Montreal, QC, Canada
Email‡: zahid.akhtar.momin@emt.inrs.ca

***Abstract*— Smartphones have become the pervasive personal computing platform. Recent years thus have witnessed exponential growth in research and development for secure and usable authentication schemes for smartphones. Several explicit (e.g., PIN-based) and/or implicit (e.g., biometrics-based) authentication methods have been designed and published in the literature. In fact, some of them have been embedded in commercial mobile products as well. However, the published studies report only the brighter side of the proposed scheme(s), e.g., higher accuracy attained by the proposed mechanism. While other associated operational issues, such as computational overhead, robustness to different environmental conditions/attacks, usability, are intentionally or unintentionally ignored. More specifically, most publicly available frameworks did not discuss or explore any other evaluation criterion, usability and environment-related measures except the accuracy under zero-effort. Thus, their baseline operations usually give a false sense of progress. This paper, therefore, presents some guidelines to researchers for designing, implementation, and evaluating smartphone user authentication methods for a positive impact on future technological developments.**

***Index Terms*—Biometrics, Smartphone Authentication, Human-Computer Interaction, Mobile Biometrics**

I. Introduction

Recent years have witnessed a lot of effort targeting the development of secure and usable authentication solutions for smartphones. Each proposed method has some pros and cons though, the published manuscripts however only report the brighter side of their solutions, i.e., a majority of papers highlight and report higher accuracy attained by their solutions, while other operational issues, e.g., power consumption, computational overhead, and usability factors, etc., are not normally mentioned. For example, the newly proposed touch-based solutions [1] have shown to be accurate in in-lab settings, however, the accuracy dropped significantly when tested in the wild [2]. Similarly, face recognition has shown to be very accurate, however, its performance is significantly affected by environmental variability. Additionally, it tends to get the user annoyed owing to the fact that the user has to take a lot of selfies throughout the day [3].

In this paper, we present some of the guidelines, particularly targeting researchers of smartphone authentication domain, for helping them in designing, implementation, and evaluation of upcoming proposed biometric-based schemes prior to publishing. The intention is to help researchers in publishing high-quality and highly impactful products that could potentially be embraced by the practitioners. To the best of our knowledge, this is the first effort towards mobile biometrics benchmark evaluation methodology. We hope that this article will grow to become a valuable tool for research in this arena. Specifically, we hope that the guidelines presented here will (i) further the pace of innovation in mobile biometrics, (ii) increase the likelihood that outcomes attained in a lab-setting would generalize to real-world operational scenarios, (iii) stimulate inter-disciplinary research and development in novel mobile biometric authentication to unleash its full potential.

All in all, the published studies in the literature consider the evaluation of the biometric systems in an independent and non-unified way, thus hard to be used universally. Therefore, this article is an attempt to devise a comprehensive evaluation methodology for smartphone biometric authentication considering various parameters such as modality attractiveness to the users.

II. Guidelines

A. Data Collection Protocol

The use of human biological data for the purpose of identity management is termed as biometric recognition or simply biometrics. Biometrics can be categorized as physical (based on the physical body parts), behavioral (based on human behaviors), chemical (based on the events that happens in the human body) and cognitive (based on brain responses). Physiological biometrics based on face, fingerprint, hand-geometry, etc., and behavioral biometrics based on voice, gait, keystroke, signature, etc., are extremely popular in the development of mobile biometrics [4].

Biometric-based user authentication for mobile devices has been studied for a long time [4]. As a result, large commercial deployments based on face, i.e., facial recognition to Complement Galaxy S8s iris scanning[1], fingerprint on iPhone[2]

[1]https://mobileidworld.com/facial-recognition-galaxy-s8-003102/
[2]https://www.theguardian.com/technology/2013/sep/22/apple-iphone-fingerprint-scanner-hacked

2017 International Carnahan Conference on Security Technology (ICCST)
978-1-5386-1585-0/17/$31.00

and iris[3] on Fujitsu mobile, already exist.

Recent trends in mobile biometric research reveal that behavioral biometric based authentication solutions have obtained significant attraction in both industry and academia. The reason behind their popularity is (i) their data can be collected unobtrusively, (ii) generally, they do not require any additional hardware, (iii) apparently they are considered as secure and (iv) the patterns can be easily revoked. Some of the popular human behaviors, e.g., gait, keystroke, touch and voice, are extensively tested and evaluated for authentication purposes. The data hence associated with these modalities is extensively available, but for the newer modalities, their associated data needs to collected.

Most published papers utilized the collected data in the lab and under supervised conditions in one-session. This strategy of data collection is cumbersome and time consuming to both supervisors and the participants, on one hand, and somehow biased since human behavior tend to vary a lot with respect to time, on the other. Hence, a concrete conclusion on such experimentation cannot be drawn.

Guideline 1: *We recommend to explain well the purpose, methodology and possible research outcome to the participants before starting the data collection process. Data should be collected anonymously or their data privacy should be ensured.*

Guideline 2: *We recommend to collect data in a natural way, i.e., data should be collected in multiple sessions so that the participant should not be able to memorize the behavior. Another possible way is to ask participants to do required actions/gestures, as and when they need to interact with their mobile devices. The participants should be given the due time (without explicitly asking them to complete the testing, in some days) for the data collection.*

Guideline 3: *Stronger claims regarding the accuracy of the proposed scheme should not be made on the obtained results of few users. We suggest to recruit as many participants as possible and cover diverse maximum population.*

Guideline 4: *We recommend to collect as many samples/templates as possible to draw a concrete conclusion.*

B. Classification Protocol

1) One-Class vs. Binary Class Classification: In biometric authentication scenarios, a specific classifier needs to be trained on a dataset $\mathcal{D}$, consisting of samples over $(\mathrm{x},\ y)$; where x is variable with a set of attributes $\mathrm{x} = \{x_1 \dots x_n\}$ and y is the ground truth label. Later, the trained classifier, based on its trained model, needs to correlate the query sample with correct label. In case of binary class classification, the classifier needs to be trained on the data of two classes and for one class classification (anomaly detection) the training data is comprised of data from just one user, i.e., the owner of the smartphone. Binary classifiers are more accurate in discriminating between the owner and non-owner because they are trained on the dataset corresponding to both classes. Whereas, the one-class classifiers need to be trained on the dataset of only one class (owner) [5], [6], and have to analyze the deviation between the query and template sample to accept/reject the owner, they are considered, comparatively, less accurate.

Guideline 5: *As mobile devices are considered very personal and sharing of the biometric samples among the users may lead to privacy breaches. The mobile user authentication problem essentially is a one-class classification problem, it is thus unreasonable to formulate mobile user authentication as the binary class classification problem.*

2) Cross-Validation vs. Training/Testing Methodology: Cross-validation is a way to evaluate performance of the classifier on a given dataset. This method ensures that the each instance is used both for training and testing the classifier. K-fold cross validation is a very common approach in the literature proposed for biometric-based user authentication on mobile devices [13]. K-fold cross-validation method randomizes the data and divides into K folds (e.g., K = 10). In each iteration, 9 out of 10 folds are used for model training, and remaining fold is used for testing the model. The process is repeated till all the folds get tested and results are averaged over all folds and final results are reported.

Training/Testing split method is also a way to evaluate the performance of the classifier. The dataset is generally split into two parts, i.e., training and testing sets. The model is trained on the training set (generally, 66% of the whole data) and the remaining test dataset is used for testing the model.

Since the number of collected observations to evaluate mobile biometric systems are generally less, the cross-validation method looks justified from machine learning perspective, however, it seems a bit unrealistic in the real world. For instance, some of the banks have signature recognition systems and they require some attempts (e.g., 5 or 6) for the classifier training and every time the customer wants to access their service, they have to provide the testing/query sample to test the classifier.

Guideline 6: *We consider classifier training with initial set of observations, e.g., first 5 or 10, more realistic as compared to using a large fraction for the classifier training.*

3) Success Metric: The published studies normally report the accuracy of the classifier (both for binary and one-class) in terms of True Acceptance Rate (TAR), False Rejection Rate (FRR), False Acceptance Rate (FAR), True Rejection Rate (TRR), Equal Error Rate (EER), Receiver Operating Characteristics Curves (ROC) and/or overall accuracy. The

[3]http://www.ibtimes.co.uk/unlocking-phone-your-eyes-fujitsu-iris-recognition-tech-coming-smartphones-2015-1490297

Related work	Input Method	Sensors	Classifiers	Users	Data Source	Results
[7]	movement	Ac, Gr, Gy, Mg, and Os	MLP & RF	53	O & I	TAR = 96%, EER = 4%
[8]	touch + movement	Ac, Os, compass and touchscreen	SVM	28	O & I	accuracy = 95.78%
[9]	touch + movement	Ac, MIC, Location and touchscreen	NaiveBayes	7	O & I	accuacy = 97%
[10]	movement	Ac, Os, Gy and Mg	n-gram language model	20	O & I	accuracy = 71.3%
[11]	tap + movement	Ac (3 variants), Os, Gr, Mag	BN & RF	12	O & I	EER = 1%
[12]	touch + movement	Ac and rotation	SVM	100	O, O & I	0% - 24.99% FAR
[6]	tap + movement	Ac, Gy and Mg	SM,SE, SVM	100	O	EER = 6.92%
[5]	touch + movement	Ac (3 variants), Os, Gr, Mag	BN, RF, KNN, MLP	30	O	FAR = 3.1%, FRR = 5.2%

TABLE I: Comparison of different authentication mechanisms. Our comparison is limited to the work which involve sensory readings and user interaction with the device, i.e., tapping, touch, etc. *denotes different user positions, namely, sitting, standing, walking, lying on the sofa, walking up & downstairs. Ac, Os, Gr, Gy, Mag stands for accelerometer, orientation, gravity, gyroscope and magnetometer, respectively. Similarly BN, RF, SM, SE, SVM, KNN, MLP stands for bayesNET, random forest, scaled manhattan, scaled euclidean, support machine classifiers, K nearest neighbor and multilayer perceptron. O denotes the smartphone owner and I denotes Impostors. O & I mean the system was trained with data from owner and impostors.

most common approach adopted to estimate the values of above-mentioned matrices is as follows: for any given dataset containing N users with n samples per user, the averaged results over N iterations are reported such that at each iteration a specific user with all its available samples is profiled as legitimate user while the rest $N-1$ users are labeled as impostors. A general overview of some recent works is presented in Table I. It is easy to see that due to the diverse use of performance matrices, one can not easily and properly compare the proposed techniques. In other words, in most of the published articles, the evaluation, unfortunately, is limited to reporting only one performance metric, thereby making it difficult (if not impossible) to draw a comprehensive conclusion.

Guideline 7: *Mobile biometric researchers should also include some other impactful measures, e.g., Failure to Acquire Rate (FTAR)*[4] *and Failure to Enroll Rate (FTER)*[5] *in their result card, beside the commonly used metrics.*

C. Usability Analysis

1) Sample Acquisition Time: It is the time required to capture a sample for authentication by the user. It is one of the most important factor, since users get annoyed by longer required acquisition time that may possibly result in complete removal of the biometric solution. Few representative published mechanisms' acquisition time can be seen in Table II.

Guideline 8: *While developing the biometric solutions, biometric researchers should minimize the required sample acquisition time in order to increase the acceptability of their proposed scheme and modality.*

[4]This failure may occur due to the inability of the system to capture the required quality sample to extract the sufficient number of features. This may occur due to several reasons, e.g., insufficient sample quality or inability to capture the trait, etc.

[5]This error rate is similar to FTAR and can be related to the systems inability to store the new reference sample.

Method	Sample Acquisition Time (s)
Hold & Sign [5]	3.5
PIN	3.7
Password	7.46
Voice	5.15
Face	5.55
Gesture	8.10
Face + Voice	7.63
Gesture + Voice	9.91

TABLE II: Sample acquisition time for different methods adapted from [5].

2) Classifier Training and Testing Times: The training and testing times are the required times by the classifier to be trained on training samples for estimation of operating parameters and to accept/reject any authentication attempt, respectively. The classifier training process is one time procedure. Therefore, user may compromise on the training time but not on the testing time. If the classifier takes longer time to make decision about the provided authentication attempt, the user may get annoyed and might possibly not use the solution any more. Some of the recent studies have reported these times as shown in the Table III.

Guideline 9: *While developing the biometric solutions, biometric researchers should take care of the testing time their proposed schemes would take to authenticate/reject the authentication attempt. Larger testing time could end up in annoying the user and won't get the wide user acceptability.*

Ref.	Classifier	Training Time	Testing Time
Hold & Sign [5]	MLP	3.5 - 9.3s	0.215 - 0.250 s
Lee et al. [14]	SVM	6.07s	20s
Li et al. [8]	Sliding patterns	n.a	0.648s
Nickel et al. [15]	KNN	90s	30s

TABLE III: Comparison of recently reported training/testing time.

3) Applicability to all users: Authentication solution providers/researchers should pay much attention to the users demographic groups. As smartphone users could be anyone

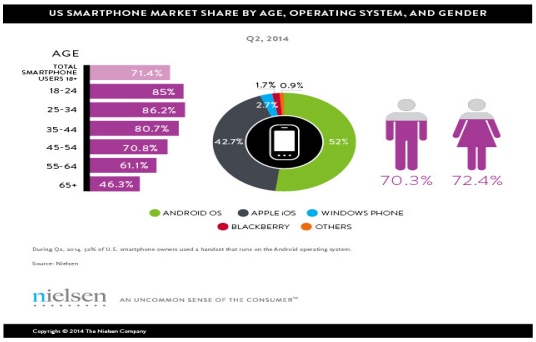

Fig. 1: Smartphone user statistics[nielsen.com]

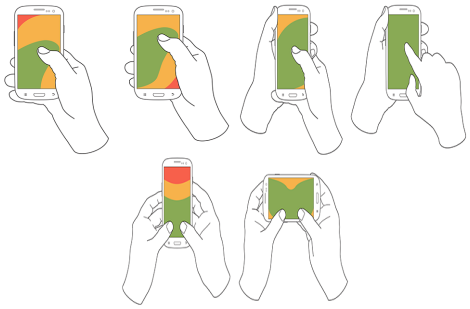

Fig. 2: Smartphone user hand preference[uxmatters.com]

among male/female (see Figures 1, teenager/old man or left/right hander. Likewise, some user also enjoys employing both hands (see Figure 2). Most of the teenagers avoid locking their smartphone because they are either unaware of the risks to their data privacy or they see it as wastage of time. Hence, the proposed authentication solution(s) should be not only attractive enough to the teenagers but also acceptable to other groups of users.

4) Applicability in different situations: Owing to their portability, advanced features and services, smartphones offer the users much more usage opportunities compared to desktops/laptops. Smartphones usability in diverse situations, environments or positions has been very popular among its users, which is another vital evaluation criteria. Needless to say, the smartphone owners may use their smartphones in different positions or situations such as sitting, standing, walking, lying on the sofa/bed, walking upstairs/downstairs, jogging, driving, and cycling, etc., Therefore, while providing input sample, they need to hold the smartphone in such a way that the maximum screen becomes visible to them. Ideally, any proposed mobile biometric authentication mechanism should be situation/positions/activity independent (e.g., fingerprint recognition). Unfortunately, majority of the proposed behavioral-based authentication solutions are limited to some specific activities and positions [11] [6] [5], hence a conclusive claim about their user acceptability cannot be made.

Guideline 10: *The newly proposed authentication scheme needs to be evaluated in multiple common activities in order to obtain a clear picture of their final accuracy..*

5) Role of Hardware/Device Variability: There are thousands of smartphone manufacturers across the globe and the number is ever increasing each year. Each well-known manufacturer, i.e., Samsung, Apple, Lenovo, Huawei and LG Electronics, usually launch multiple smartphone models every year. Besides hardwares, these smartphones also differ in their operating systems (e.g., iOS, Android, Windows, blackberry). It is very common to observe that the accuracy and attack-resistance of the mechanism does not remain consistent over different smartphones. In particular, the performance accuracy largely varies among the device models as well as manufacturers due to differences in hardware and software.

Guideline 11: *It would be worth investigating to evaluate the newly proposed authentication scheme on different devices and/or multiple models and reporting the results accordingly.*

6) Software Usability Scale (SUS): The Software Usability Scale (SUS)[6]- a 10-questions based assessment tool has widely been used to record users experience with the system, in general. Users' response to each of the question is recorded on a five-point scale ranging from "Strongly Disagree" to "Strongly Agree". The computed score is a value between 0 and 100 where a higher score indicates higher usability. A raw SUS score can be transformed to a percentile [16] or to a grading scale [17], allowing easier interpretation of results.

Guideline 12: *Research proposing new mobile biometric should also include initial usability evaluation to get an impression of their user acceptability.*

D. Performance Analysis

Any proposed authentication mechanism should be lightweight, rather than resource hungry, in order to attain wider usability and acceptability. For instance, they should not involve any extra overhead on CPU and memory. Likewise, they should be computationally inexpensive both in training and decision-making. There are two kinds of approaches to accurately determine the power consumption of mobile applications: hardware-based and software-based. Though hardware approaches are highly accurate [6], they are expensive. Software-based techniques are easier to use and are mostly available for free. Examples of power profiling tools are powerTutor, Trepn, and Gsam battery monitor. It will hugely help to further the state-of-the-art if all proposed mechanism would also evaluate and report properly the power consumption estimated by either way.

[6]http://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html.

1) Power Overhead: The amount of time a smartphone can run with a single charge is termed as 'battery life'. The battery life of smartphones depends on the amount of usage, therefore it varies from person to person. For instance, an addicted smartphone user could have less battery life and vice-versa. Battery life is one of the most important parameters that increases the usability while enhancing user's experience of smartphones and security. Smartphones are the most personal portable-device being used frequently in everyday life. A better understanding of *where* and *how* the energy is/could be consumed may help researchers to design more acceptable authentication solutions. The users might not have any issue with extra battery consumption by the hardware (GPRS, screen, etc.). However, we can not state the same for the authentication solutions.

Guideline 13: *Optimal management of battery consumptions for these devices are imperative, which also means that any proposed mobile biometric recognition system must not consume much power to be adopted in the real-world applications at large scale.*

2) Computational Overhead: The Central Processing Unit (CPU) is the heart of any smartphone that process data (and also executes the instructions). When user processes plenty of data at once, the data (or large part of it) gets loaded into the RAM to be used later by the CPU. The CPU processes the data and potentially more memory is occupied owing to the fact that the CPU at times keeps the processed information in the memory. Although, RAM, and CPU do not have to correlate, but they often do. Biometric authentication solutions may end up utilizing CPU and RAM extensively leading thus to bad user-experience. For example, a biometric solution that is $\approx 100\%$ secure, but if it halts the smartphone for $10 - 20s$ during the authentication phase then it will be unacceptable to the user. An indication of these parameters can be estimated by CPU and memory profilers such as Trepn[7], CPU Monitor[8], which are available on the Google play.

The best way to evaluate the impact of proposed authentication solution is to test it with the benchmark applications. These benchmark applications execute certain usage scenarios and evaluate its impact on CPU, memory usage, and I/O, etc. The well-known benchmarks available are AnTuTu, GreekBench, Quadrant Standard, and Vellamo Mobile, which can be found on the Google play.

Guideline 14: *It is strongly recommended to report CPU and memory overhead usage estimation for the proposed mechanism(s) to avoid any bad user-experience.*

[7]https://play.google.com/store/apps/details?id=com.quicinc.trepn&hl=en

[8]https://play.google.com/store/apps/details?id=com.bigbro.ProcessProfiler&hl=en

E. Adversarial Analysis

It is very common trend to report only the performance accuracy of the proposed mobile biometric authentication, while ignoring the security analysis against attacks. It would greatly help to position the proposed system among the existing similar mechanisms, if each study would report both accuracy as well as robustness against various attacks. This evaluation approach is different from the recognition approach, since a certain identity is intentionally targeted by attackers/impostors to fool the system [18]. This kind of attack is also known as 'non-zero-effort' attack. While, in recognition approach impostor normally does not intentionally aims to fool the system without targeting any specific user. The potential attacks in mobile biometrics can be categorized as *random attacks*, *mimic attacks*, and *engineered attacks*. In the following, we discuss these attacks in detail:

1) Random Attacks: Lets us consider a scenario, where a genuine user lost or forgot their smartphone, somewhere. The person who finds the phone, consequently, that person would attempt to unlock the smartphone illicitly by pretending as the genuine user. We call this attack as random because the phone finder can only use random tries (without knowing the details of the implemented mechanism and the legit user's behavior) to unlock the smartphone. Random attacks have a great practical relevance because the scenario in which they might occur is most likely to happen in our daily lives. Moreover, they don't require advanced technical skills and, therefore the potential number of attackers is very large.

Guideline 15: *To obtain such attacks samples, the participants should be asked to try randomly unlocking the device without knowing the implemented authentication mechanism.*

2) Mimic Attacks: Lets us consider a scenario, where a genuine user lost or forgot their smartphone in common places like in the office, in canteen, in a gathering with friends, etc. The person who finds the phone is aware of the employed authentication mechanism. In order to understand this kind of attacks, during data collection or actual operation, the test adversaries should be asked to mimic the genuine user's pattern that has been provided them beforehand. For instance, in a keystroke authentication based solutions if the genuine user has adopted 1234 pin, all other remaining users should be provided this information to evaluate the classifier performance under random attacks. Similarly, the "Hold & Sign" [5] method could be evaluated under adversaries access by showing the target signature (on the smartphone screen or printed on paper, etc.) to them. Also, each adversary should be allowed sufficient attempts to fool the mechanism.

Guideline 16: *To obtain such attacks samples, a genuine user could be asked to use the mechanism in front of the test-adversaries as many times as possible. In this*

way the test-adversaries may get a better overview of the implemented mechanism as well as legitimate user's behaviors that is to be mimicked.

3) Engineered Attacks: Engineered attacks are also referred as spoofing attacks [18]. The behavioral mobile biometrics spoofing requires technical knowledge and/or resources, therefore also named as engineered attacks. Unlike physical biometrics, spoofing behavioral mobile biometrics are considered very difficult, which need a lot of time and efforts. However, recently a study conducted on spoofing touch biometrics showed the possibility [19]. The authors studied the vulnerability of the touch gestures in terms of zero-effort (where an attacker does not need to make any effort to spoof the system, namely impostor) and spoofing via a robotic device. In particular, they demonstrated how a robotic device can pose a major threat to touch-based user authentication systems. An EER of 0.035% and 0.13% were attained using support vector machine and KNN classifiers, respectively, and these EER increased up to 900% under robotic attacks.

In an another attempt [20], authors evaluated the vulnerability of touch biometrics against sophisticated adversaries. The sophisticated adversary attacks were procured by two types of methods: a population-statistics-driven attack method and a user-tailored attack method. The population-statistics-driven attacks are based on patterns gleaned from a large population of users while the user-tailored attack is based on samples stolen from the victims. Both attacks are launched by a Lego robot, which was trained on how to swipe on the touch screen. They observed an increase in FAR up to 5-times under the attacks compared to the standard zero-effort impostor attacks. They observed that the attacks increases the systems mean FAR by up to 5-times relative to the mean FAR seen under the standard zero-effort impostor attacks.

Guideline 17: *We admit that executing this type of attack is a bit time taking, cumbersome, and tricky, but the claims regarding the robustness of their proposed schemes should only be made after such evaluation.*

III. Conclusion

Mobile biometrics has been an active area of research in recent years. Physical biometrics due to their inherent security and usability limitations have become less preferred by the users. Hence the focus of the research in this domain has been shifted towards developing novel behavioral biometric-based solutions. In order to maximize the impact and usability of the proposed schemes/modalities, it becomes extremely important to evaluate comprehensively the upcoming schemes/modalities with diverse criterion. To this end, we have provided some guidelines for mobile security researchers with the intention to help them in designing, implementation, and evaluation of their schemes, which may lead to high-quality and highly impactful products.

Acknowledgment

The work was partially supported by the EIT Digital project: Android App Reputation Service (ARTS), by the European Training Network for CyberSecurity (NeCS) grant number 675320 and by the project XProbes funded by the Provincia Autonoma di Trento.

References

[1] T. Feng, Z. Liu, K.-A. Kwon, W. Shi, B. Carbunar, Y. Jiang, and N. Nguyen, "Continuous mobile authentication using touchscreen gestures," in *Homeland Security (HST), 2012 IEEE Conf. on Technologies for*. IEEE, 2012, pp. 451–456.

[2] T. Feng, J. Yang, Z. Yan, E. M. Tapia, and W. Shi, "Tips: Context-aware implicit user identification using touch screen in uncontrolled environments," in *Proceedings of the 15th Workshop on Mobile Computing Systems and Applications*. ACM, 2014, p. 9.

[3] A. De Luca, A. Hang, E. Von Zezschwitz, and H. Hussmann, "I feel like i'm taking selfies all day!: Towards understanding biometric authentication on smartphones," in *Proc. 33rd Annual ACM Conf. on Human Factors in Computing Systems*, 2015, pp. 1411–1414.

[4] W. Meng, D. S. Wong, S. Furnell, and J. Zhou, "Surveying the development of biometric user authentication on mobile phones," *IEEE Comm. Surveys & Tutorials*, vol. 17, no. 3, pp. 1268–1293, 2015.

[5] A. Buriro, B. Crispo, F. Delfrari, and K. Wrona, "Hold and sign: A novel behavioral biometrics for smartphone user authentication," in *IEEE Workshops S&P, 2016*, 2016, pp. 276–285.

[6] Z. Sitova, J. Sedenka, Q. Yang, G. Peng, G. Zhou, P. Gasti, and K. Balagani, "Hmog: A new biometric modality for continuous authentication of smartphone users," *arXiv preprint arXiv:1501.01199*, 2015.

[7] A. Buriro, B. Crispo, and Y. Zhauniarovich, "Please hold on: Unobtrusive user authentication using smartphone's built-in sensors," in *Identity, Security and Behavior Analysis (ISBA), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1–8.

[8] L. Li, X. Zhao, and G. Xue, "Unobservable re-authentication for smartphones." in *NDSS*, 2013.

[9] W. Shi, J. Yang, Y. Jiang, F. Yang, and Y. Xiong, "Senguard: Passive user identification on smartphones using multiple sensors," in *Wireless and Mobile Computing, Networking and Communications (WiMob), 2011 IEEE 7th Int'l Conf. on*. IEEE, 2011, pp. 141–148.

[10] J. Zhu, P. Wu, X. Wang, and J. Zhang, "Sensec: Mobile security through passive sensing," in *Computing, Networking and Communications (ICNC), 2013 Int'l Conf. on*. IEEE, 2013, pp. 1128–1133.

[11] A. Buriro, B. Crispo, F. Del Frari, and K. S. Wrona, "Touchstroke: Smartphone user authentication based on touch-typing biometrics." in *ICIAP Workshops*, 2015, pp. 27–34.

[12] 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*. ACM, 2013, pp. 187–190.

[13] A. Buriro, "Behavioral biometrics for smartphone user authentication," Ph.D. dissertation, University of Trento, 2017.

[14] W.-H. Lee and R. B. Lee, "Multi-sensor authentication to improve smartphone security," in *Conf. on Info. Sys. Security and Privacy*, 2015.

[15] C. Nickel, T. Wirtl, and C. Busch, "Authentication of smartphone users based on the way they walk using k-nn algorithm," in *8th International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*. IEEE, 2012, pp. 16–20.

[16] J. Sauro. (2011) Measuring usability with the system usability scale (sus). [Online]. Available: http://www.measuringu.com/sus.php

[17] A. Bangor, P. T. Kortum, and J. T. Miller, "An empirical evaluation of the system usability scale," *Intl. Journal of Human–Computer Interaction*, vol. 24, no. 6, pp. 574–594, 2008.

[18] Z. Akhtar, C. Micheloni, and G. L. Foresti, "Biometric liveness detection: Challenges and research opportunities," *IEEE Security Privacy*, vol. 13, no. 5, pp. 63–72, Sept 2015.

[19] 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*. ACM, 2013, pp. 599–610.

[20] 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 (TISSEC)*, vol. 18, no. 4, p. 14, 2016.