MULTIMODAL SMARTPHONE USER AUTHENTICATION USING TOUCHSTROKE, PHONE-MOVEMENT AND FACE PATTERNS

Zahid Akhtar[1], *Attaullah Buriro*[2], *Bruno Crispo*[2], *Tiago H. Falk*[1]

[1]INRS-EMT, University of Quebec, Canada
[2]University of Trento, Italy

ABSTRACT

Recently, researches have shown to employ implicit behavioral biometrics via built-in sensors (e.g., gyroscope) for user identification on smartphones. The majority of prior studies are based on unimodal systems, which suffer from low accuracy and spoofing. In this paper, we present an unconstrained and implicit multimodal biometric system for smartphones using touchstroke, phone-movement and face patterns. The proposed framework authenticates the user by taking *silently* into account micro-movements of the phone[1], movements of the user's finger during typing on the touchscreen, and user's face features. We also collected a mobile multimodal dataset of touchstroke and phone-movement patterns in the wild from 95 subjects. Preliminary experimental analysis on accuracy and usability show promising results.

Index Terms**—** Mobile Biometrics, Authentication, Multimodal Biometrics, Touch Dynamics, Behavioral biometrics.

1. INTRODUCTION

Nowadays, smartphones have become pervasive personal computing platform [1]. In fact, they have become personal assistants, which are widely being used not only for basic communications but also as a tool to store e-mail, personal photos, Internet banking and payments, and social networking. Thus, one of the biggest concerns is security and privacy leakage. Explicit authentication methods, e.g., PIN, password, graphical pattern, are the most common security strategy available on commercial smartphones [2]. They are susceptible to guessing or smudge attacks, besides being user-unfriendly and time-consuming [3]. Biometric-based user authentication on smartphones have been recently adopted. For instance, iPhone 6s, Android KitKat mobile OS, and Fujitsu's Arrows NX F-04G employ fingerprint, face and iris to recognize individuals, respectively. Nevertheless, they still face unresolved security[2] and usability issues [4, 5]. The average smartphone user checks their device 150 times per day[3]. If explicit passcode/biometric-based authentication takes 2 seconds, the typical user spends 5 minutes unlocking their device/app every day. In fact, recent studies have shown that about 34%, 47% and 40% users did not use passcode, fingerprint or any kind of authentication mechanisms, respectively, on their mobile devices citing low usability [4, 6]. Moreover, fingerprint scanner in smartphones most likely fails when it encounters dry fingers [7].

Researchers have lately proposed behavioral biometrics-based solutions for smartphone user authentication using built-in sensors such as accelerometer, gyroscope, and microphones. Some advantages of these techniques are: minimal user interaction, unobtrusive data collection, and no additional hardware required [8]. The authors in [9] proposed a mobile framework *Sensec* based on sequence of actions that collects accelerometer, orientation sensor, touch screen, and magnetometer data for user authentication. While, Buriro *et al.* [10] employed a sensor enhanced touchstroke mechanism for smartphone biometric system, which analyzes how a person holds her phone and how one types her 4-digit free-text PIN. A specific five-finger touch gestures based user recognition system for Apple iPad was presented in [11]. However, this method is not feasible for small touchscreens of typical smartphones. Blanco-Gonzalo *et al.* [12] studied handwritten signature recognition on mobile device. A systematic analysis of prior methods show that they principally either require high user cooperation, work under constrained environment and protocol, use single modality that can be easily spoofed and mimicked (e.g., signature) [13] or do not yet exhibit low enough error rates. Moreover, no multimodal database using built-in three-dimensional sensors (e.g., accelerometer, gravity sensor, and magnetometer) is publicly available. Similarly, despite several advantages the state-of-the-art in unconstrained and unobtrusive multimodal smartphone biometric authentication is relatively new.

This paper presents an unconstrained and implicit multimodal biometric system for smartphones user identification based on face, user's hand micro-movements and touchstroke patterns. Specifically, the system captures *silently* the micro-

[1] They are low-frequency and low-amplitude unnoticeable hand movements caused by touchstroke actions.

[2] Just two days after the iPhone5s hit the market, it was fooled by a fingerprint spoof [13].

[3] http://abcnews.go.com/blogs/technology/2013/05/cellphone-users-check-phones-150xday-and-other-internet-fun-facts/

movement of phone, user's face, and timing of touch-typing[4] while she is entering a text-independent PIN/password to verify the user's identity. To the best of our knowledge, this is the first work that explores combination of face, phone-movement and touchstroke for smartphone user recognition. The presented system does not require user to remember any password, graphical pattern or gesture, since in each attempt user can enter novel text-independent PIN of fixed length. Moreover, typing a PIN is easier than writing something on touchscreen (e.g., signature/pattern). Even if impostor knows what is being entered as a PIN, access will still be denied because the system employs keystroke dynamics features (that is specific to each individual) and not the PIN itself, plus impostor cannot mimic the phone micro-movement of genuine user. Additionally, acquiring a person's face unobtrusively using phone's frontal camera is reasonably easier than acquiring other biometric traits, e.g., fingerprint, palmprint and iris. Also, use of multi modality makes the proposed system more secure than uni-modal ones, as spoofing only one or two modalities would not suffice to grant access to the phone [14].

To demonstrate the efficacy of the proposed system, we also collected a multimodal dataset of touchstroke and phone-movement patterns from 95 subjects with multiple smartphones under uncontrolled fashion where user could sit, stand or walk. Experimental analysis shows promising results with more practical accuracy and usability. Namely, this study shows that multimodal biometrics can be integrated with smartphones in a user-friendly manner with significant improved usability and security.

The paper is organized as follows. Section 2 describes the proposed multimodal smartphone user authentication. Experimental protocol, dataset, results and analysis are discussed in Section 3. Conclusions are drawn in Section 4.

2. PROPOSED MULTIMODAL AUTHENTICATION SYSTEM

Smartphones are now being used intensively for accessing and storing sensitive personal data, thereby making user authentication an issue of paramount importance. The balance between security and usability however is a challenging task. For instance, according to the survey result reported in [15], the iris and voice biometric traits were ranked highest in terms of perceived security protection, but lowest in terms of usability mainly due to high requirement of user cooperation. Recently, it has been shown [8, 9, 16] that each user holds, interacts and moves her phone in a unique way, which can also be utilized for implicit user authentication. In this work, we propose an unconstrained and implicit multimodal biometric user authentication system based on user's hand micro-movements, touchstroke, and face patterns.

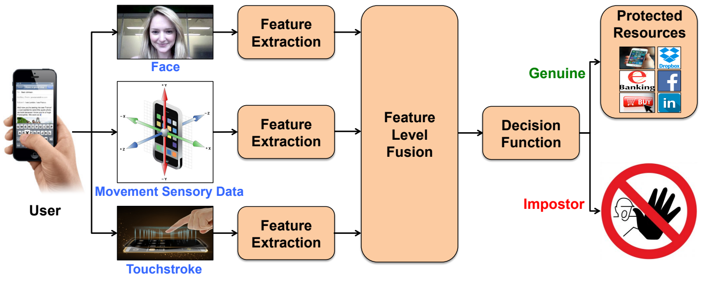

Fig. 1. Proposed smartphone user authentication framework.

The proposed framework, as illustrated also in Fig. 1, profiles *silently* the user's hand micro-movements caused by touch and face captured from frontal camera of the smartphone[5] simultaneously with touchstrokes while user is entering 8-digit text-independent passcode to access either the phone or any app. The hand micro-movements pattern analyzed is of a very short period of time, i.e., time taken by the user to enter 8 digits. The rationale behind choosing this time duration is because it was empirically determined that this time is sufficient enough for pattern discrimination and this duration is too short for an adversary to debug the device [17]. The 8-digit passcode was adopted owing to the fact that most apps, document storing/sharing and social networking services use minimum 8-digit password.

In particular, the captured three mobile biometric traits (i.e., face, hand micro-movement and touchstroke) are separately processed to extract respective salient features. These three independent features are combined using a feature level fusion scheme (here, feature concatenation). The resultant feature vector is then fed to a classifier to obtain final binary decision: genuine or impostor. If the user is classified as a genuine user, the system does not interrupt the owner's interactions with the smartphone. Otherwise, the system alerts the owner of the phone (e.g., sending an email), and may stealthily isolate the impostor from accessing sensitive functionality [18], or ask for explicit authentication [20, 21]. Since the proposed method does not use components of a specific mobile platform, thus it can be implemented for any mobile operating system.

In the proposed system, face processing sub-system is composed of two steps: face localization and face feature extraction. Face localization is achieved by the classification and regression tree analysis (CART) based on Haar features [22]. The detected face is first normalized, and then Binarized Statistical Image Features (BSIF) technique is applied to obtain the face features. BSIF is a local image descriptor that represents each pixel as a binary code obtained by computing its response to a set of filters, which are learned from natural images using an ICA (Independent Component Analysis) based unsupervised scheme. Finally, the histogram of the pix-

[4] Touch-typing is the act of typing input on the touchscreen of a smartphone.

[5] Since the system captures user's face unobtrusively, it thus assumes that during authentication process face of the user is in the range of *angle of view*.

els' binary string values allows one to characterize the texture properties within the face sub-regions.

For phone movements and touchstroke dynamics, the method leverages three built-in three-dimensional sensors: the accelerometer, the gravity sensor, and the magnetometer. By applying the Low-Pass and the High-Pass filters[6] with the parameter $\alpha = 0.5$ to the accelerometer data, two additional sensory readings are also obtained; we call them LPF and HPF values. The Raw, LPF, and HPF accelerometer readings produce gravity, apparent transient forces acting on the phone caused by the user activity, and exact acceleration applied by the user on the phone values, respectively. While, gravity sensor and magnetometer provide the magnitude and direction of the gravity force applied on the phone, and strength and/or direction of the magnetic field in three dimensions, respectively. All these sensors generate a continuous stream in x, y and z dimensions. We also added a fourth dimension to all sensors, and named it *magnitude* that is represented as:

$$S_M = \sqrt{(a_x^2 + a_y^2 + a_z^2)} \tag{1}$$

where S_M is the resultant dimension, and a_x, a_y and a_z are the accelerations along the X, Y and Z directions. The touchstroke features are extracted using typing *n-graph*, namely dwell time and flight time from each typing pattern.

For final classification whether the user is genuine or impostor, Multilayer Perceptron (MLP) and Random Forest (RF) classifiers were used because of their simplicity, fast computation and resistance against over-fitting.

3. EXPERIMENTS

Here, we provide an experimental evaluation of the proposed multimodal smartphone user authentication framework.

3.1. Dataset

MOBIO Face database [23]. This public dataset consists of face samples collected from 150 subjects using a NOKIA N93i mobile phone under realistic and uncontrolled environment. It was captured in phase I and II, which respectively include 21 and 11 videos per user with 6 sessions. Fig. 2 shows examples from the MOBIO database.

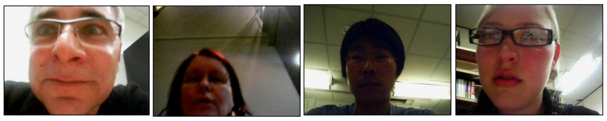

Fig. 2. Examples from MOBIO face database.

Movement and Touchstroke database. We developed a customized Android application for the purpose of this data collection. The app collects data from multiple sensors (i.e., accelerometer, gravity, orientation, magnetometer and gyroscope sensors) and LPF and HPF values [19]. While Android supports data collection in four fixed delays (often termed as *Sensor_Delay_Modes*): *SENSOR_DELAY_NORMAL*, *SENSOR_DELAY_GAME*, *SENSOR_DELAY_UI* and *SENSOR_DELAY_FASTEST* with fixed delay of 0.2, 0.02, 0.06 and $0s$, respectively. We observed *SENSOR_DELAY_NORMAL* to be too slow to sense the movements, and *SENSOR_DELAY_UI* and *SENSOR_DELAY_FASTEST* very likely to include the noise during data collection. Thus, we used *SENSOR_DELAY_GAME* in this paper. The data was collected from 95 users under uncontrolled environment using a crowd sourcing platform. The users were required to install the app and enter 8-digit touch-types in different activities, i.e., *sitting*, *standing*, and *walking*. The dataset was collected in three phases (days); 30 samples in each activity per phase from each user. This database is made publicly available by the authors.

3.2. Experimental Protocol

Since no multi-modal data sets including face, phone micro-movements and touckstroke are publicly available, we created a chimerical multi-modal data sets by associating the three modalities of pairs of clients of the available *MOBIO* and *Movement and Touchstroke* data sets without regard to age and/or gender. Note that building chimerical data sets is a widely used approach in experimental investigations on multimodal biometrics [24]. Following steps were performed on the detected face: conversion to grayscale, histogram equalization, normalization to 175×175, and feature extraction using BSIF filter of size 11×11 with a length of 12 bits. From every 3-dimensional sensor, 4 data streams were collected to extract 4 statistical features (mean, standard deviation, skewness, and kurtosis) in each data stream. Data from every sensor were then transformed into a 4×4 features matrix. In total 16 features from all four dimensions of each sensor were obtained. Likewise, 14 features were extracted based on touch-typing timing from the text-independent 8-digit passcode entered by the user. The two adopted classifiers (i.e., MLP and RF) were used from the WEKA workbench [25] for user authentication. The performance was evaluated using TAR (True Acceptance Rate) and EER (Equal Error Rate), which respectively are the fraction of legitimate user attempts classified correctly and where false rejection and false acceptance become equal. To mitigate the class imbalance problem, we used randomly selected n ($= 5$ or 10) and all samples of a genuine user and an impostor, respectively, as training set, whereas remaining samples of the genuine user and all samples of an impostor (that was not used in training) were used as testing set; the testing phase was repeated 94 times using different impostor in each run. Furthermore, all the above procedure was repeated 94 times for the given user using a different impostor in training in each run. Reported results are average values over the $94 \times 94 \times 95$ runs.

[6] http://developer.android.com/guide/topics/sensors/sensors_motion.html

Activity	Classifiers	Movements		Touch		Face		Movements+Touch		Movements+Face		Touch+Face		Movements+Touch+Face	
		TAR	EER	TAR	EER	TAR	EER	TAR	EER	TAR	EER	TAR	EER	TAR	EER
Sitting	MLP	0.98	0.02	0.78	0.22	0.57	0.43	0.97	0.03	0.93	0.07	0.70	0.30	0.92	0.08
	RF	0.97	0.03	0.82	0.18	0.52	0.48	0.97	0.03	0.94	0.06	0.60	0.40	0.95	0.05
Standing	MLP	0.98	0.02	0.80	0.20	0.54	0.46	0.98	0.02	0.94	0.06	0.72	0.28	0.94	0.06
	RF	0.98	0.02	0.84	0.16	0.49	0.51	0.98	0.02	0.97	0.03	0.73	0.27	0.97	0.03
Walking	MLP	0.98	0.02	0.81	0.19	0.58	0.42	0.98	0.02	0.94	0.06	0.76	0.24	0.94	0.06
	RF	0.97	0.03	0.84	0.16	0.52	0.58	0.98	0.02	0.96	0.04	0.78	0.22	0.97	0.03

Table 1. Performance of both classifiers (averaged over all 95 users) under *sitting, standing,* and *walking* activities for unimodal, bimodal and trimodal scenarios based on 5 training samples.

Activity	Classifiers	Movements		Touch		Face		Movements+Touch		Movements+Face		Touch+Face		Movements+Touch+Face	
		TAR	EER	TAR	EER	TAR	EER	TAR	EER	TAR	EER	TAR	EER	TAR	EER
Sitting	MLP	0.99	0.1	0.79	0.21	0.70	0.30	0.99	0.01	0.97	0.03	0.79	0.21	0.96	0.04
	RF	0.99	0.1	0.86	0.14	0.54	0.46	0.99	0.01	0.97	0.03	0.70	0.30	0.97	0.03
Standing	MLP	0.99	0.01	0.82	0.18	0.61	0.39	0.99	0.01	0.99	0.01	0.82	0.18	0.99	0.01
	RF	0.99	0.01	0.87	0.13	0.50	0.50	0.99	0.01	0.99	0.01	0.80	0.20	0.99	0.01
Walking	MLP	0.99	0.01	0.84	0.16	0.66	0.34	0.99	0.01	0.98	0.02	0.82	0.18	0.98	0.02
	RF	0.99	0.01	0.91	0.09	0.54	0.56	0.99	0.01	0.98	0.02	0.80	0.20	0.99	0.01

Table 2. Performance of both classifiers (averaged over all 95 users) under *sitting, standing,* and *walking* activities for unimodal, bimodal and trimodal scenarios based on 10 training samples.

3.3. Experimental Results

The experimental results of smartphone user authentication are presented in Tables 1 and 2 with 5 and 10 training samples, respectively, under three conditions: *sitting*, *standing* and *walking*. Several observations may be extracted from Tables 1–2: i) the proposed framework presents a high potential as a simple, unconstrained, implicit and novel method for multimodal smartphone user authentication, which reaches a great recognition accuracy for diverse practical in-the-wild conditions (e.g., *sitting*, *standing*, and *walking*) without requiring user's co-operation or one to remember any PIN/pattern; ii) RF classifier and phone micro-movement modality, by and large, outperform other considered classifier and modalities, respectively, owing to RF's ability of reducing the variances, averaging out the biases and most unlikeliness of overfitting, and micro-movement modality's very low intra- and high inter-class variation; iii) increasing the number of training samples increases identification accuracy, it however reduces the usability as in practice most users likely train their systems with fewer samples, e.g., during our data collection 49.5% users said they would prefer to use less than or maximum 5 samples for training the system; iv) among adopted modalities face comparatively achieved high error rates mainly because samples in-the-wild generally exhibit low texture clarity (see Fig. 2), plus BSIF hugely depends on ICA that works better for non-Gaussian data, while we observed that face dataset samples in this study tend strongly to be Gaussian; v) phone micro-movement is the most accurate modality among three, however unimodal systems are vulnerable to spoofing. In fact, it has been shown in [14] that spoofing the best individual modality creates serious security breaches, i.e., attacker has higher chances to evade the system when one spoofs most accurate modality and vice versa; vi) integration of two/three modalities is not consistently beneficial. Generally, benefits of fusion are exploited when modalities show complementary nature [7]. However, since in this case face modality is significantly worse than others, the performance of multimodal system is thus below the best performing modality. Nonetheless, the performance drop in trimodal system is not very much. Moreover, it is worth noticing that the proposed multimodal system (Phone movement+Touchstroke+Face) still procures higher accuracy using only 5 or 10 training samples than multimodal methods in [26] and [5] using 25 and 30 samples, respectively. All in all, fusion improves security against spoofing, since an attacker has lower chances to evade multimodal systems (even if he spoofs all fused traits) than to evade each single unimodal system [14]. Namely, spoofing three modalities simultaneously is much harder than single modality.

4. CONCLUSION

This paper presents an unconstrained and implicit smartphone multimodal biometric system using touchstroke, phone-movement and face patterns. The proposed architecture identifies the user by taking *silently* into account micro-movements of phone, movements of user's finger during typing on touchscreen, and user's face features. This study also collected and shares publicly a mobile multimodal dataset of touchstroke and phone-movement patterns in the wild from 95 subjects. Experimental results show high accuracy with increased *security* and *usability*. Since proposed method is generic it can be implement on any smartphone to unlock the device and/or a separate authentication service for various applications, e.g., mobile banking. Future work will focus on analyzing vulnerability of the proposed system against spoofing and thereby developing anti-spoofing techniques.

5. REFERENCES

[1] W. Liu, T. Mei and Y. Zhang, “Instant Mobile Video Search With Layered Audio-Video Indexing and Progressive Transmission”, *IEEE Transactions on Multimedia*, 16(8):2242–2255, 2014.

[2] W. Meng, D. S. Wong, S. Furnell and J. Zhou, “Surveying the Development of Biometric User Authentication on Mobile Phones”, *IEEE Communications Surveys & Tutorials*, 17(3):1268–1293, 2015.

[3] F. Schaub, R. Deyhle, M. Weber, “Password entry usability and shoulder surfing susceptibility on different smartphone platforms”, *Proc. of the 11th Int'l Conf. on Mobile and Ubiquitous Multimedia*, pp.1–10, 2012.

[4] A. De Luca, A. Hang, E. von Zezschwitz, H. Hussmann, “I Feel Like I'm Taking Selfies All Day!: Towards Understanding Biometric Authentication on Smartphones”, *ACM Conference on Human Factors in Computing Systems*, pp. 1411–1414, 2015.

[5] A. Buriro, B. Crispo, F. Delfrari, K. Wrona, “Hold and Sign: A Novel Behavioral Biometrics for Smartphone User Authentication”, *IEEE Security and Privacy Workshops*, pp. 276–285, 2016.

[6] M. Harbach et al., “It's a Hard Lock Life: A Field Study of Smartphone (Un)Locking Behavior and Risk Perception”, *Symposium On Usable Privacy and Security*, pp. 213–230, 2014.

[7] D. Maltoni, M. Maio, A.K. Jain, S. Prabhakar, “Handbook of Fingerprint Recognition”, *Springer*, 2009.

[8] Z. Sitov et al., “HMOG: New Behavioral Biometric Features for Continuous Authentication of Smartphone Users”, *IEEE Transactions on Information Forensics and Security*, 11(5):877–892, 2016.

[9] J. Zhu, P. Wu, X. Wang, J. Zhang, “Sensec: Mobile security through passive sensing”, *IEEE Int'l Conf. on Comp., Networking and Comm. (ICNC)*, pp. 1128–1133, 2013.

[10] A. Buriro, B. Crispo, F. DelFrari, and K. Wrona, “Touchstroke: Smartphone user authentication based on touch-typing biometrics”, *New Trends in Image Analysis and Processing–ICIAP Workshops*, pp. 27–34, 2015.

[11] N. Sae-Bae, K. Ahmed, K. Isbister, N. Memon, “Biometric-rich gestures: a novel approach to authentication on multi-touch devices”, *Proc. of Conf. on Human Factors in Computing Systems*, pp. 977–986, 2012.

[12] R. Blanco-Gonzalo, R. Sanchez-Reillo, O. Miguel-Hurtado, J. Liu-Jimenez, “Performance evaluation of handwritten signature recognition in mobile environments”, *IET biometrics*, 3(3):139–146, 2014.

[13] Z. Akhtar, C. Micheloni and G. L. Foresti, “Biometric Liveness Detection: Challenges and Research Opportunities”, *IEEE Security & Privacy*, 13(5):63–72, 2015.

[14] Z. Akhtar, “Security of Multimodal Biometric Systems against Spoof Attacks”, PhD thesis, *University of Cagliari*, Italy, 2012.

[15] P.S. Teh, N. Zhang, A.B. Jin Teoh, K. Chen, “A survey on touch dynamics authentication in mobile devices”, *Computers & Security*, 59(2):210–235, 2016.

[16] A. Buriro, B. Crispo, F. Del Frari, J. Klardie, K. Wrona, “ITSME: Multi-modal and unobtrusive behavioural user authentication for smartphones”, *International Conference on Passwords*, pp. 45–61, 2014.

[17] T. Vidas, D. Votipka, N. Christin, “All Your Droid Are Belong to Us: A Survey of Current Android Attacks”, *USENIX Conf. on Offensive Technologie.*, pp. 1–10, 2011.

[18] W. Shi et al., “SenGuard: Passive User Identification on Smartphones Using Multiple Sensors”, *Proc. of IEEE Int'l Conf. on Wireless and Mobile Computing, Networking and Communications*, pp. 141–148, 2011.

[19] A. Buriro, “Behavioral Biometrics for Smartphone User Authentication”, PhD thesis, *University of Trento, Italy*, 2017.

[20] O. Riva et al., “Progressive Authentication: Deciding when to Authenticate on Mobile Phones”, *Proc. of USENIX Conf. on Security Symposium*, pp. 1–15, 2012.

[21] E. Hayashi, S. Das, S. Shahriyar, J. Hong, I. Oakley, “CASA: Context-aware Scalable Authentication”, *Symposium on Usable Privacy and Security*, pp. 1–10, 2013.

[22] R. Lienhart et al., “Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection”, *Symp. on Pattern Recognition*. pp. 297–304, 2003.

[23] C. McCool et al., “Bi-Modal Person Recognition on a Mobile Phone: using mobile phone data”, *IEEE Workshop on Hot Topics in Mobile Mutlimedia*, pp. 1–6, 2012.

[24] A. Ross, K. Nandakumar, A.K. Jain, “Handbook of Multibiometrics”, *Springer*, 2006.

[25] G. Holmes et al., “Weka: A machine learning workbench”, *Proc. Australian and New Zealand Conference on Intelligent Information Systems*, pp. 357–361, 1994.

[26] M. Shahzad, A. X. Liu, A. Samuel, “Behavior Based Human Authentication on Touch Screen Devices Using Gestures and Signatures”, *IEEE Transactions on Mobile Computing*, 15(12):1–14, 2016.