Privacy-preserving Biometric-driven Data for Student Identity Management: Challenges and Approaches

Christos Fidas
Department of Electrical and Computer Engineering, University of Patras, Greece
fidas@upatras.gr

Marios Belk
Cognitive UX GmbH, Germany
belk@cognitiveux.de

David Portugal
Institute of Systems and Robotics, University of Coimbra, Portugal
davidbsp@isr.uc.pt

Andreas Pitsillides
Department of Computer Science, University of Cyprus, Cyprus
andreas.pitsillides@ucy.ac.cy

ABSTRACT

Biometric technologies are being considered lately for student identity management in Higher Education Institutions, as they provide several advantages over the traditional knowledge-based and token-based authentication methods, *i.e.*, biometrics provide high security entropies, convenience and a sense of technological modernity to the end-users. While biometric technologies have many benefits from both a security and usability point of view, still there is a need for innovative user identity management solutions that continuously identify and authenticate students during academic and teaching activities. In addition, biometrics entail several threats and weaknesses with regards to the privacy of data stored about the user, which negatively affect the user acceptance and the wider adoption of biometrics due to regulatory and legal issues. In this paper, we refer to our ongoing research on intelligent and continuous online student identity management for improving security and trust in European Higher Education Institutions. We further highlight based on the literature, existing challenges, threats and state-of-the-art approaches with regards to preserving the privacy of biometric-driven data.

CCS CONCEPTS

• **Security and privacy** → Human and societal aspects of security and privacy; Privacy protections; Human and societal aspects of security and privacy; Usability in security and privacy.

KEYWORDS

User Authentication, Biometrics, Privacy, Security, Blockchain

ACM Reference Format:
Christos Fidas, Marios Belk, David Portugal, and Andreas Pitsillides. 2021. Privacy-preserving Biometric-driven Data for Student Identity Management: Challenges and Approaches. In *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '21 Adjunct), June 21–25, 2021, Utrecht, Netherlands.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3450614.3464470

UMAP '21 Adjunct, June 21–25, 2021, Utrecht, Netherlands

ACM ISBN 978-1-4503-8367-7/21/06.
https://doi.org/10.1145/3450614.3464470

1 INTRODUCTION

User identity management is a critical aspect of any information system today aiming to assure that end-users have the appropriate access to sensitive data and services. Core components of user identity management relate to: *i) user authentication* aiming to validate that the end-users are allowed to access the system by requiring them to provide various authentication factors, or a combination of them (*e.g.*, textual and graphical passwords, push notifications on smartphones, Time-based One Time Passwords (TOTP), graphical Transaction Authentication Numbers (PhotoTAN), biometrics, etc.) [1, 2, 3]; *ii) continuous user authentication and identification* aiming to verify the end-user's identity in real-time (after successfully authenticating), while carrying out tasks [4, 5]; and *iii) access control* aiming to regulate user access to the system resources [6].

In this context, biometric-based authentication within user identity management represents a significant and evolving field of research and practice [9], as it entails several benefits from both a security and usability point of view. Specifically, biometrics can create high entropies of the secret biometric data used for authentication, minimize administration expenses, offer convenience to end-users compared to traditional knowledge-based (*e.g.*, passwords) and token-based (*e.g.*, TOTP) solutions, and they provide a sense of technological modernity to the end-users [10, 11]. Common approaches for biometric-based authentication are based on the end-users' physical (*e.g.*, fingerprint, iris, face, voice, etc.) and/or behavioral characteristics (*e.g.*, typing patterns, interaction patterns, engagement patterns, etc.) [7, 8]. Such technologies have become an important means for enforcing strict security policies in a variety of domains such as government, education, etc. [9, 10, 12].

2 BIOMETRICS IN THE HIGHER EDUCATION DOMAIN

Higher Education Institutions (HEIs) already started considering the adoption of biometric technology for seamlessly identifying and/or authenticating students within teaching and learning activities, academic services, etc. The need for such solutions has become even more evident due to the new COVID-19 pandemic, in which

most HEIs adopted a fast transition towards a completely online academic, teaching and learning paradigm. This transition embraces severe challenges related to deploying trustworthy and credible user identity management solutions aiming to enforce academic integrity in online examinations, course attendance, minimize administration expenses, as well as offer convenience to the end-users. However, nowadays, user authentication in most online learning platforms is compromised only by the user login task based on a single-point of entry (*i.e.*, typically through a traditional password system). As such, it remains questionable whether the user who logged in, is in reality the one who attends or fulfills the educational activities.

One direction to address the aforementioned issues is to design, develop and evaluate an integrated and multi-tier framework for student identity management by utilizing privacy-preserving biometric-based solutions for face-, voice- and interaction-based continuous user authentication and/or identification techniques for higher education. For face-based identification, one approach would be to process real-time video data collected through low-cost Web-based camera devices. This can be achieved using diverse image processing and computer vision techniques for face recognition, based on either classical approaches like eigenfaces [13], or recently popular deep learning approaches [14, 15]. Specifically, face biometrics would be performed through face recognition and analysis, combined with eye gaze behavior biometrics by analyzing the end-users' eye gaze data and visual behavior during interaction (*e.g.*, [16, 17]). For voice-based identification, voice data would be collected through built-in microphone devices aiming to identify users in real-time by extracting features via state-of-the-art signal processing techniques (*e.g.*, [18, 19, 20]). Finally, interaction-based analysis techniques based on keystroke dynamics, interaction behavior, etc. (*e.g.*, [21]) would identify and authenticate the end-users based on their interaction behavior in real-time.

Such an endeavor entails several challenges with regards to preserving the privacy of the biometric data stored about the end-users. Such challenges and state-of-the-art mitigation solutions are described next.

3 PRIVACY-PRESERVING CHALLENGES OF BIOMETRICS AND STATE-OF-THE-ART APPROACHES

Core threats and challenges for designing secure and privacy-preserving biometric technologies, as discussed in [7, 8, 11, 25], relate to: *i) security of biometric data:* in several cases biometric data are not secret (*e.g.*, fingerprints can be extracted from surfaces that the user touched, faces can be easily acquired from public online sources, voices can be recorded, etc.) [8]; *ii) privacy of biometric data:* biometric data that are stored could expose sensitive information about the end-users (*e.g.*, ethnic origin, health information, etc.) [11]; and *iii) revocability of biometric data:* in case biometric data are compromised, it is nearly impossible to revoke the biometric data of the end-users [8].

Hence, there is a strong need to implement and deploy innovative solutions aiming to assure that biometric data are processed and stored in such a way to achieve high levels of security and sustain privacy-preservation aspects. To this end, state-of-the-art approaches (see [7, 8, 11, 24, 25] for a detailed analysis of such approaches) for preserving the privacy of biometric-driven data include: *i) biometric templates*, which are digital representations of certain features extracted from a biometric sample (*e.g.*, the shape of a user's hand), without storing the exact raw biometric data of the user to avoid potential privacy issues in case of a compromised data set; *ii) biometric encryption* techniques have been employed to address privacy issues in biometrics. Given the high variability of biometric data, traditional cryptographic hashing approaches may not be suitable for biometric data [11], hence, different cryptographic tools have been applied such as homomorphic encryption (see [11, 25] for a detailed analysis of such cryptographic tools); *iii) protocol-based approaches* have been proposed in order to protect the privacy of biometric data, *e.g.*, secure multiparty computation protocol, zero-knowledge proof protocol, etc. [25]; and *iv) blockchain technology* has specific features that can address several of the existing challenges in privacy-preserving biometrics, *i.e.*, its distributed nature addresses the problem of single-point of failure, elimination of third-parties and potential privacy leakage, monitoring and access to trustable and unmodifiable history logs [6, 22, 23, 24].

4 CONCLUSIONS

In this paper, we overview existing challenges and commonly adopted approaches for privacy-preserving biometric-driven data for user identity management. Specifically, in the context of HEIs, the COVID-19 global pandemic catalyzed the need for continuous student authentication during online activities, such as exams, and laboratory or practical classes. Ongoing research on face-, voice- and interaction-based authentication pave the way for the globalized adoption of reliable automatic authentication techniques, but also impose important privacy, security, ethical and user experience concerns. Our future work on these topics aims to provide innovative and credible identity management methods for continuously identifying students during online learning activities, by applying a User-Centered Design (UCD) methodology to promote and build new knowledge grounded on evidence-based research, through the implementation of several end-user studies with different Higher Education Institutions in Europe.

ACKNOWLEDGMENTS

This work is partially supported by a new European project, TRUSTID - Intelligent and Continuous Online Student Identity Management for Improving Security and Trust in European Higher Education Institutions (Grant Agreement No: 2020-1-EL01-KA226-HE-094869), which is funded by the European Commission within the Erasmus+ 2020 Programme.

REFERENCES

[1] Mare, S., Baker, M., Gummeson, J. (2016). A Study of Authentication in Daily Life. In Proc. SOUPS 2016, 189-206
[2] Ometov, A., Bezzateev, S., Maekitalo, N., Andreev, S., Mikkonen, T., Koucheryavy, Y. (2020). Multi-Factor Authentication: A Survey. Cryptography, 2(1), 1
[3] Constantinides, A., Fidas, C., Belk, M., Pietron, A., Han, T., Pitsillides, A. (2021). From Hot-spots towards Experience-spots: Leveraging on Users' Sociocultural Experiences to Enhance Security in Cued-recall Graphical Authentication. International Journal of Human-Computer Studies, 149

[4] Gonzalez-Manzano, L., De Fuentes, J., Ribagorda, A. (2019). Leveraging User-related Internet of Things for Continuous Authentication: A Survey. ACM Computing Surveys, 52(3), article 53, 38 pages
[5] Buschek, D., De Luca, A., Alt, F. (2015). Improving Accuracy, Applicability and Usability of Keystroke Biometrics on Mobile Touchscreen Devices. In Proc. ACM CHI 2015, 1393-1402
[6] Rouhani, S., Deters, R. (2019). Blockchain Based Access Control Systems: State of the Art and Challenges. ACM Web Intelligence, 423-428
[7] Jain, A.K., Nandakumar, K., Ross, A. (2016). 50 Years of Biometric Research: Accomplishments, Challenges, and Opportunities. Pattern Recognition Letters, 79, 80-105
[8] Rui, Z., Yan, Z. (2018). A Survey on Biometric Authentication: Toward Secure and Privacy-Preserving Identification. IEEE Access, 7, 5994-6009
[9] Bhalla, A. (2020). The Latest Evolution of Biometrics. Biometric Technology Today, 2020 (8), 5-8
[10] Gray, S.L. (2017). Biometrics in Schools: The Role of Authentic and Inauthentic Social Transactions. BSA Conference 2017
[11] Pagnin, E., Mitrokotsa, A. (2017). Privacy-Preserving Biometric Authentication: Challenges and Directions. Security and Communication Networks, 2017, 7129505
[12] Labati, R.D., Genovese, A., Muñoz, E., Piuri, V., Scotti, F., Sforza, G. (2016). Biometric Recognition in Automated Border Control: A Survey. ACM Computing Surveys, 49(2), Article 24, 39 pages
[13] Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J. (1997). Eigenfaces vs. Fisherfaces: Recognition using Class Specific Linear Projection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7), 711-720
[14] Guo, G., Zhang, N. (2019). A Survey on Deep Learning based Face Recognition. Computer Vision and Image Understanding, 189, 102805
[15] Schroff, F., Kalenichenko, D., Philbin, J. (2015). Facenet: A Unified Embedding for Face Recognition and Clustering. IEEE Conference on Computer Vision and Pattern Recognition, 815-823
[16] Constantinides, A., Belk, M., Fidas, C., Pitsillides, A. (2020). An Eye Gaze-driven Metric for Estimating the Strength of Graphical Passwords based on Image Hotspots. ACM Intelligent User Interfaces (IUI 2020), ACM Press, 33-37
[17] Constantinides, A., Belk, M., Fidas, C., Pitsillides, A. (2019). On the Accuracy of Eye Gaze-driven Classifiers for Predicting Image Content Familiarity in Graphical Passwords. ACM User Modeling, Adaptation and Personalization (UMAP 2019), ACM Press, 245-249
[18] Boles, A., Rad, P. (2017). Voice Biometrics: Deep Learning-based Voiceprint Authentication System. System of Systems Engineering Conference, IEEE, 1-6
[19] Nagrani, A., Chung, J.S., Xie, W., Zisserman, A. (2020). Voxceleb: Large-scale Speaker Verification in the Wild. Computer Speech & Language, 60, 101027
[20] Ravanelli, M., Bengio, Y. (2018). Speaker Recognition from Raw Waveform with Sincnet. IEEE Spoken Language Technology Workshop, IEEE, 1021-1028
[21] Belk, M., Portugal, D., Christodoulou, E., Samaras, G. (2015). Cognimouse: On Detecting Users' Task Completion Difficulty through Computer Mouse Interaction. ACM Conference Extended Abstracts on Human Factors in Computing Systems, 1019-1024
[22] Zhang, R., Xue, R., Liu, L. (2019). Security and Privacy on Blockchain. ACM Computing Surveys, 52(3), article 51
[23] Tran, Q.N., Turnbull, B.P., Wu, H., de Silva, A., Kormusheva, K., Hu, J, (2021). A Survey on Privacy-Preserving Blockchain Systems (PPBS) and a Novel PPBS-Based Framework for Smart Agriculture. IEEE Open Journal of the Computer Society, 2, 72-84
[24] Sarier, N.D. (2018). Privacy Preserving Biometric Identification on the Bitcoin Blockchain. International Symposium on Cyberspace Safety and Security (CSS 2018), 254-269
[25] Tran, Q.N., Turnbull, B.P., Hu, J. (2021). Biometrics and Privacy-Preservation: How Do They Evolve? IEEE Open Journal of the Computer Society, 2, 179-191