User Authentication through Keystroke Dynamics by means of Model Checking: A Proposal

Fabio Di Tommaso*, Michele Guerra*, Fabio Martinelli†,
Francesco Mercaldo† *, Massimo Piedimonte*, Giovanni Rosa*, Antonella Santone*
*Department of Biosciences and Territory, University of Molise, Pesche (IS), Italy
{f.ditommaso2, m.guerra4, m.piedimonte, g.rosa1}@studenti.unimol.it
{francesco.mercaldo, antonella.santone}@unimol.it
†Institute for Informatics and Telematics, National Research Council of Italy (CNR), Pisa, Italy
{fabio.martinelli, francesco.mercaldo}@iit.cnr.it

***Abstract*—The current authentication systems based on password and pin code are not enough to guarantee attacks from malicious users. For this reason, in the last years, several studies are proposed with the aim to identify the users basing on their typing dynamics. In this paper, we propose the adoption of formal methods to discriminate between different users by exploiting a set of keystroke features. The idea behind the proposed method is to identify the users silently and continuously during their typing on a monitored system. To perform such user identification effectively, we consider a feature vector able to capture the typing style that is specific to each given user. By considering this feature model, in detail we propose to consider model checking with logic temporal properties to discriminate between different users using a set of keystroke features.**

I. Introduction

The study of the keystroke dynamics is largely diffused in the last years allowing to identify or authenticate individuals basing on the way they type on a keyboard: the timing of each keystroke, the pressure applied when typing and some additional specific features for the mobile devices (i.e., the orientation of the device, accelerometer, the size of touch and the location of touch) [1], [2]. The reason for this interest is that keystroke analysis [3], [4] should improve the existing computer security systems providing ad additional security layer in performing a more efficient user identification with respect to the traditional authentication approaches that are recognized to be vulnerable [3]. Looking, for instance, to the password based authentications, such methods have been repeatedly proven to be easy to compromise and are usually limited to serve to facilitate a one-off authentication judgment at the start of a session [3] (some systems should require seamless and continuous monitoring). However, the keystroke analysis can cooperate with the traditional authentication systems allowing to perform a more robust identification. An advantage of keystroke analysis for user authentication relies on being not intrusive by design: the stream of typed keys generates an event flow that can be analyzed with several techniques in real-time and without interfering with user activity or behavior. Furthermore, keystroke analysis is well supported by keylogger software that can be easily used to capture, collect and extract the typing events, along with the relevant data, with very reduced costs. In last years, several machine learning approaches have been considered for the keystroke analysis task [1].

Differently from the current state-of-the-art, based on the adoption of artificial intelligence for keystroke analysis, in this paper we propose a method for granting the user access exploiting model checking. Model checking represents, for a given finite-state model of a system, the exhaustively and automatically checking whether this model meets a given specification [5]. To solve such a problem algorithmically, both the model of the system and the specification are formulated in some precise mathematical language. To this end, the problem is formulated as a task in logic, namely to check whether a given structure satisfies a given logical formula [6]. This general concept applies to many kinds of logic and suitable structures [7], [8]. A simple model checking problem is verifying whether a given formula in the propositional logic is satisfied by a given structure [9], [10].

II. Related Works

The study of the keystroke dynamics is largely diffused in the last years allowing to identify or authenticate individuals basing on the way they type on a keyboard [1], [11]. The reason for this interest is that keystroke analysis should improve the existing computer security systems providing ad additional security layer in performing a more efficient user identification with respect to the traditional authentication approaches that are recognized to be vulnerable [3].

A huge amount of algorithms are used in the past three decades to perform user identification basing on the analysis of keystroke dynamics [12]. A vast amount of these user identification algorithms are based on the adoption of statistical machine learning and neural network techniques. Looking to the neural networks approaches authors in [1] investigate the use of a behavioral biometric providing the evidence that probabilistic neural network can be more effective and have higher classification accuracy with respect to a typical back-propagation trained neural-network.

Artificial neural networks are also applied in [13]–[15] where authors respectively use perceptron, back-propagation neural network, and Art-2 neural network to perform user classification. The study of these approaches shows that they

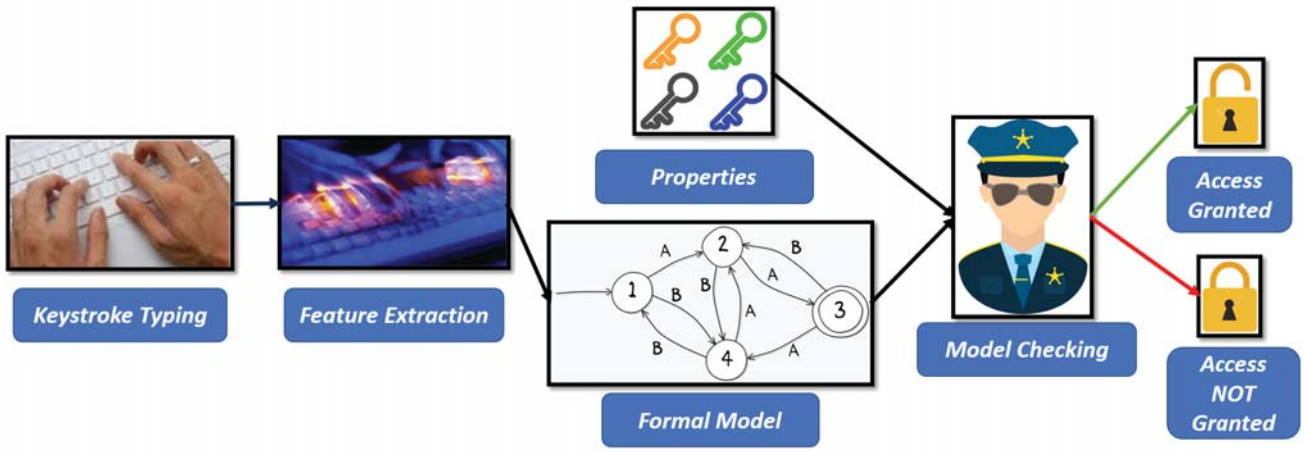

Fig. 1: The proposed approach for user authentication.

mainly suffer from a slow speed of the training model. Moreover, they are characterized by hand selection of the model architecture and tuning of parameters. Finally, these models have poor generalization capabilities [16]. Statistical machine-learning approaches use K-nearest neighbors classifiers [17], Bayesian classifiers [18] and support vector machines [19] to perform the keystroke classification. Support vector machines algorithm shows good results (highly efficiency) in identification and verification tasks. The above-described studies show good performances but their evaluation is limited to a set of their own features extracted from their own datasets. This limitation is discussed in [20] where authors introduce a keystroke dynamics benchmark dataset.

Finally, a very limited number of studies [2] propose deep learning approaches to increase the keystroke analysis approaches performances. Deep learning techniques are known to be more suitable to handle large intra-class variations and noisy biometric data.

As emerged from the state-of-the-art discussion, this proposal represents the first attempt to authenticate users silently and continuously by means of model checking technique.

III. Applying Model Checking to Keystroke Dynamics

The proposed approach for user authentication relies on being not intrusive by design: as a matter of fact from the stream of typed keys an automaton is generated and analysed in real-time without interfering with user activity. The underlying assumption behind the approach is that users can be identified on the base of their typing dynamics.

Figure 1 shows the proposed approach for user authentication through keystroke dynamics.

The method is comprising of following phases:

- free-text keystroke typing is gathered exploiting, for instance, a daemon able to retrieve the typed text;
- a set of features is computed from the keystroke typing;
- the feature set is translated in processes (i.e. a formal model) by exploiting the Calculus of Communicating Systems [21];
- a set of properties in mu-calculus logic [22] describing the user behaviour is defined;
- we invoke the formal verification environment (for instance, the CWB-NC[1]) to verify if the user behaviour properties satisfy the automaton generated from the user keystroke typing;
- whether the model checker outputs *true* the access is granted to the user under analysis, otherwise the access is denied.

In detail we propose the adoption of following feature set, coherently with previous studies [1], [3], [11] on user authentication by exploiting keystroke dynamics:

- *key up*: it represents the key press event;
- *key down*: it represents the key release event;
- *Error Corrections*: it refers to the frequency of Backspace or Delete (DEL) keypress events during the typing (i.e. deletes per minute);
- *Inter Keys Interval*: it is the range (in milliseconds) between two consecutive *key down* events;
- *Inter Keys Interval*: it consists in the range (in milliseconds) between the last key press event and the middle keypress event of the last trigram of events;
- *Inter Keys Interval*: It is the range (in milliseconds) between the last KeyDown event and the first key event in the previous trigram of events;
- *Words per minute*: number of words per minute, evaluated for each typed sentence.

IV. Conclusion and Future Work

The keystroke analysis for user identification is a very diffused task in the last years. The importance of the topic

[1]https://www3.cs.stonybrook.edu/~cwb/

is linked to the need of new users identification approaches given the vulnerability and the limitations of the current authentication systems usually based on password and pin codes (single authentication for the entire session). The existing approaches are mainly based on machine learning classifiers and are evaluated on small and ad-hoc datasets.

We propose the adoption of model checking for the user authentication task. In detail starting from of a set of features, gathered by keystroke typing, a formal model is generated. Thus, adopting a formal verification environment, we check whether on this model a set of properties related to user behaviours are satisfied. As future work, we plan to implement the proposed approach and to perform a real-world large scale evaluation aimed to demonstrate the effectiveness and the portability (by considering, for instance, PC and mobile environment) of the following proposal.

ACKNOWLEDGMENTS

This work has been partially supported by MIUR - SecureOpenNets and EU SPARTA and CyberSANE projects and WEBREPUTO POR FESR 2014-202

REFERENCES

[1] K. Revett, F. Gorunescu, M. Gorunescu, M. Ene, S. T. d. Magalhaes, and H. M. D. Santos, "A machine learning approach to keystroke dynamics based user authentication," *Int. J. Electron. Secur. Digit. Forensic*, vol. 1, no. 1, pp. 55–70, May 2007. [Online]. Available: http://dx.doi.org/10.1504/IJESDF.2007.013592

[2] E. Hellström, "Feature learning with deep neural networks for keystroke biometrics: A study of supervised pre-training and autoencoders," *Dissertation.*, 2018.

[3] P. S. Dowland, S. M. Furnell, and M. Papadaki, *Keystroke Analysis as a Method of Advanced User Authentication and Response*. Boston, MA: Springer US, 2002, pp. 215–226. [Online]. Available: https://doi.org/10.1007/978-0-387-35586-3_17

[4] F. Bergadano, D. Gunetti, and C. Picardi, "User authentication through keystroke dynamics," *ACM Trans. Inf. Syst. Secur.*, vol. 5, no. 4, pp. 367–397, Nov. 2002. [Online]. Available: http://doi.acm.org/10.1145/581271.581272

[5] L. Brunese, F. Mercaldo, A. Reginelli, and A. Santone, "Formal methods for prostate cancer gleason score and treatment prediction using radiomic biomarkers," *Magnetic resonance imaging*, 2019.

[6] A. Santone and G. Vaglini, "Abstract reduction in directed model checking ccs processes," *Acta informatica*, vol. 49, no. 5, pp. 313–341, 2012.

[7] L. Brunese, F. Mercaldo, A. Reginelli, and A. Santone, "Radiomic features for medical images tamper detection by equivalence checking," *Procedia Computer Science*, vol. 159, pp. 1795–1802, 2019.

[8] A. Santone, "Automatic verification of concurrent systems using a formula-based compositional approach," *Acta Informatica*, vol. 38, no. 8, pp. 531–564, 2002.

[9] F. Mercaldo, V. Nardone, A. Santone, and C. A. Visaggio, "Hey malware, i can find you!" in *Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), 2016 IEEE 25th International Conference on*. IEEE, 2016, pp. 261–262.

[10] A. Cimitile, F. Martinelli, F. Mercaldo, V. Nardone, and A. Santone, "Formal methods meet mobile code obfuscation identification of code reordering technique," in *2017 IEEE 26th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*. IEEE, 2017, pp. 263–268.

[11] M. L. Bernardi, M. Cimitile, F. Martinelli, and F. Mercaldo, "Keystroke analysis for user identification using deep neural networks," in *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019, pp. 1–8.

[12] A. B. J. T. Pin Shen Teh and S. Yue, "A survey of keystroke dynamics biometrics," *The Scientific World Journal*, 2013.

[13] R. Banerjee, S. Feng, J. S. Kang, and Y. Choi, "Keystroke patterns as prosody in digital writings: A case study with deceptive reviews and essays," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, October 2014, pp. 1469–1473. [Online]. Available: http://www.aclweb.org/anthology/D14-1155

[14] S. Haider, A. Abbas, and A. K. Zaidi, "A multi-technique approach for user identification through keystroke dynamics," in *Smc 2000 conference proceedings. 2000 ieee international conference on systems, man and cybernetics. 'cybernetics evolving to systems, humans, organizations, and their complex interactions' (cat. no.0*, vol. 2, Oct 2000, pp. 1336–1341 vol.2.

[15] C. C. Loy, W. K. Lai, and C. P. Lim, "Keystroke patterns classification using the artmap-fd neural network," in *Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP 2007)*, vol. 1, Nov 2007, pp. 61–64.

[16] Y. Deng and Y. Zhong, "Keystroke dynamics user authentication based on gaussian mixture model and deep belief nets," in *ISRN Signal Processing*, vol. 2013, 2013.

[17] V. Shanmugapriya and G. Padmavathi, "Keystroke dynamics authentication using neural network approaches," in *Information and Communication Technologies*, V. V. Das and R. Vijaykumar, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 686–690.

[18] F. Monrose and A. D. Rubin, "Keystroke dynamics as a biometric for authentication," *Future Gener. Comput. Syst.*, vol. 16, no. 4, pp. 351–359, Feb. 2000. [Online]. Available: http://dx.doi.org/10.1016/S0167-739X(99)00059-X

[19] P. Kang, S.-s. Hwang, and S. Cho, "Continual retraining of keystroke dynamics based authenticator," in *Advances in Biometrics*, S.-W. Lee and S. Z. Li, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 1203–1211.

[20] R. A. Maxion and K. S. Killourhy, "Keystroke biometrics with number-pad input," in *2010 IEEE/IFIP International Conference on Dependable Systems Networks (DSN)*, June 2010, pp. 201–210.

[21] R. Milner, *Communication and concurrency*, ser. PHI Series in computer science. Prentice Hall, 1989.

[22] C. Stirling, "An introduction to modal and temporal logics for ccs," in *Concurrency: Theory, Language, And Architecture*, ser. LNCS, A. Yonezawa and T. Ito, Eds., vol. 491. Springer, 1989, pp. 2–20.