




Extended Abstract

# Mouse Behavior Analysis Based on Artificial Intelligence as a Second-Phase Authentication System <sup>†</sup>

Daniel Garabato <sup>\*</sup>, Jorge Rodríguez García, Francisco J. Novoa  and Carlos Dafonte 

Centro de investigación CITIC, Universidade da Coruña, Campus de Elviña s/n, 15071 A Coruña, Spain

<sup>\*</sup> Correspondence: daniel.garabato@udc.es

<sup>†</sup> Presented at the II XoveTIC Congress, A Coruña, Spain, 5–6 September 2019.

Published: 1 August 2019



**Abstract:** Nowadays, a wide variety of computer systems use authentication protocols based on several factors in order to enhance security. In this work, the viability of a second-phase authentication scheme based on users' mouse behavior is analyzed by means of classical Artificial Intelligence techniques, such as the Support Vector Machines or Multi-Layer Perceptrons. Such methods were found to perform particularly well, demonstrating the feasibility of mouse behavior analytics as a second-phase authentication mechanism. In addition, in the current stage of the experiments, the classification techniques were found to be very stable for the extracted features.

**Keywords:** authentication; computer security; artificial intelligence

## 1. Introduction

Computer security has been one of the main issues and areas of interest and research since the origin of Information Technology. A wide variety of security measures have been developed over the years in order to try to protect different type of assets, from physical systems to software services and data [1]. Authentication systems can be considered a key component of these security models, becoming the basis for access control systems and the first barrier against threats. Nowadays, almost every computer system or service relies on an authentication process in order to check the identity of the users and to determine the access permissions that should be granted according to the users' specific profiles.

Such an authentication process can be performed by means of different methods or factors [2]. The most common ones are based on knowledge, such as a password or a personal identification number, or some device or item that the user owns, such as an identity card or a bank card. More sophisticated methods based on users' biometric properties, such as fingerprint, retina or voice, have evolved over the last years, so that they can verify users' identity in a very precise and robust manner. These methods have become more and more popular along with the use of smartphones and tablets, which use them to grant access to the device features. In addition, behavioral biometrics have arisen from the analysis of users' behavior during their interaction with computer systems as an alternative way of authentication. Different behavioral features, such as keyboard typing and mouse usage, have been widely studied as possible authentication factors [3], but their high variability over time make them unstable to become a primarily authentication method.

During the last years, classical authentication schemes have moved towards a two-step verification process in order to strengthen authentication. Under this scheme, users must prove their identity in two different and complementary ways, typically a password-based method is used as primary verification and, then, a generated code which is sent to the user's phone is also required to successfully complete the process.

The second verification step could be extended so that it is periodically repeated all along the users' session to verify their identities in a transparent manner, so that they are not continuously bothered and required to identify themselves. To this purpose, behavioral authentication methods could be useful as they do not require any specific device and they just monitor users' activity. Therefore, in this work we conduct an analysis of a mouse-based authentication mechanism in order to be used as a continuous second step verification process. Such a method aims to create a specific profile for each user by means of Artificial Intelligence (AI) techniques, and it requires the application of Big Data and Data Mining techniques in order to handle the enormous volume of data gathered from the users' interaction with the computer environment.

## 2. Methods

In order to conduct this experiment, mouse movement data were collected from a variety of test cases oriented to frequent user interactions with a computer, such as closing a window, forced waiting periods, or even locating the mouse pointer on the screen. To this purpose, an ad-hoc application was developed to guide the user across these test cases, while the mouse movement data were being collected in a transparent manner to the user. Due to the high variability of behavioral biometry, the entire process was repeated over three different sessions for each one of the twenty test volunteers, resulting in more than one million records of data.

Users' identity must be verified in order to make the authentication process feasible. A number of features related to different spatial-temporal measurements on mouse usage were originally proposed in [4], and they were used in this work. Since the users' behavior may significantly vary due to particular circumstances (i.e. emotional state or distractions), outlier measurements were filtered out, and then the data was scaled to the interval [0, 1] prior to any further processing. Due to the temporal and sequential basis in which the events are captured, two different data scenarios were proposed for analysis: in the first one, data were grouped in windows of 100, 200, and 500 events, respectively; and, in the second one, data were grouped in time windows of one, two and five seconds, respectively.

The analysis conducted in this work addresses the authentication process by means of two different widely used AI classification algorithms: Support Vector Machines [5] (SVM) and Multi-Layer Perceptrons [6] (MLP). User-specific profiles were built upon the three data scenarios proposed, searching for an appropriate model hyperparameterization using a grid search procedure, which also takes cross-validation into account, so that the generalization of the trained models can be assessed.

## 3. Results

The performance of the proposed experiments was measured in order to analyze whether it is possible to authenticate users according to their behavior when they are interacting with computers by means of mouse devices (Table 1). To this purpose, four different common metrics have been computed: precision, recall,  $F_1$ -score, and AUC-ROC [7].

Table 1. Average and standard deviation performance.

		Event-Based Windows			Time-Based Windows		
		100 events	200 events	500 events	1 second	2 seconds	5 seconds
MLP	Precision	0.82 ( $\pm 0.06$ )	0.83 ( $\pm 0.07$ )	0.85 ( $\pm 0.09$ )	0.79 ( $\pm 0.07$ )	0.80 ( $\pm 0.08$ )	0.80 ( $\pm 0.09$ )
	Recall	0.91 ( $\pm 0.05$ )	0.89 ( $\pm 0.07$ )	0.87 ( $\pm 0.10$ )	0.86 ( $\pm 0.07$ )	0.85 ( $\pm 0.07$ )	0.82 ( $\pm 0.10$ )
	$F_1$ -score	0.86 ( $\pm 0.05$ )	0.86 ( $\pm 0.06$ )	0.85 ( $\pm 0.08$ )	0.82 ( $\pm 0.06$ )	0.82 ( $\pm 0.06$ )	0.81 ( $\pm 0.08$ )
	AUC-ROC	0.90 ( $\pm 0.04$ )	0.91 ( $\pm 0.05$ )	0.91 ( $\pm 0.06$ )	0.87 ( $\pm 0.06$ )	0.88 ( $\pm 0.06$ )	0.87 ( $\pm 0.08$ )
SVM	Precision	0.80 ( $\pm 0.06$ )	0.81 ( $\pm 0.07$ )	0.83 ( $\pm 0.09$ )	0.78 ( $\pm 0.08$ )	0.79 ( $\pm 0.08$ )	0.79 ( $\pm 0.09$ )
	Recall	0.94 ( $\pm 0.04$ )	0.94 ( $\pm 0.06$ )	0.93 ( $\pm 0.07$ )	0.89 ( $\pm 0.06$ )	0.90 ( $\pm 0.06$ )	0.88 ( $\pm 0.08$ )
	$F_1$ -score	0.86 ( $\pm 0.04$ )	0.87 ( $\pm 0.05$ )	0.87 ( $\pm 0.07$ )	0.83 ( $\pm 0.06$ )	0.84 ( $\pm 0.06$ )	0.83 ( $\pm 0.07$ )
	AUC-ROC	0.91 ( $\pm 0.04$ )	0.92 ( $\pm 0.04$ )	0.92 ( $\pm 0.06$ )	0.88 ( $\pm 0.06$ )	0.88 ( $\pm 0.06$ )	0.89 ( $\pm 0.07$ )

The overall performance is similar for all the experiments conducted, obtaining for most users values around 0.84 and 0.89 for  $F_1$ -score and AUC-ROC, respectively. On the one hand, it was found that the usage of time windows performed slightly worse than using a fixed number of events for each chunk. The underlying cause may be the variable number of records available for each user: some of them impulsively move the mouse, whereas others use the mouse in a smoother manner. On the other hand, MLPs achieved a higher precision rate, so that it should be favored for authentication purposes rather than SVMs.

#### 4. Conclusions and Future Work

Through the analysis conducted in this work, it has been demonstrated that users' mouse-based behavior can be used for authentication purposes. Although the obtained performance may not be suitable for a single-primary authentication mechanism, it could be perfectly used as a second-phase authentication method that continuously monitors users' activity seeking after behavior anomalies, similarly to an intrusion detection system. In case of a security issue detection, the authentication monitor could decide to close users' sessions, or to report such an issue to the system administrators for further investigation.

On the basis of these results, a mouse behavior monitoring system could be developed and tested under a real environment, so that the stability of these models over time could be assessed. In addition, further analysis based on different approaches or techniques, such as deep learning, could be conducted to improve the overall performance.

**Funding:** The manpower of this work was funded by the Spanish MECD FPU16/03827, and CITIC-ACATIA contract.

**Acknowledgments:** This work was also supported by the Xunta de Galicia (Potencial Crecemento ED431B 2018/42) and the European Union (European Social Fund – ESF); we also used IT infrastructure that was acquired through the RTI2018-095076-B-C22 and ESP2016-80079-C2-2-R projects, financed by the Spanish Ministry of Science, Innovation and Universities and the Ministry of Economy, Industry and Competitiveness.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

#### References

1. Stallings, W.; Brown, L. *Computer Security: Principles and Practice*, 2nd ed.; Pearson Education Limited: London, UK, 2012.
2. Barkadehi, M.H.; Nilashi, M.; Ibrahim, O.; Fardi, A.Z.; Samad, S. Authentication systems: A literature review and classification. *Telemat. Inform.* **2018**, *35*, 1491–1511.
3. Bhatnagar, M.; Jain, R.K.; Khairnar, N.S. A Survey on Behavioral Biometric Techniques: Mouse vs Keyboard Dynamics. *IJCA* **2013**, *975*, 8887.
4. Sayed, B.; Traoré, I.; Woungang, I.; Obaidat, M.S. Biometric Authentication Using Mouse Gesture Dynamics. *IEEE Syst. J.* **2013**, *7*, 262–274.
5. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297.
6. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning Internal Representation by Error Propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*; MIT Press: Cambridge, MA, USA, 1986; Volume 1.
7. Tharwat, A. Classification assessment methods. *Appl. Comput. Inform.* **2018**. doi:10.1016/j.aci.2018.08.003.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).