International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8
***Article Received:** 20 May 2023 **Revised:** 28 May 2023 **Accepted:** 16 August 2023*

A Review on Ai based Data Authentication by Monitoring Behavioural Pattern

Praveen Kumar Chandapeta , Dr. Ajay R. Raundale
(Research Scholar Dr. APJ Abdul Kalam University,Indore,India)
(Research Supervisor Dr. APJ Abdul Kalam University,Indore,India)
pkc06032013@gmail.com*1

Abstract:
In this study, we do an experiment to examine the viability of a continually authenticating approach based on the monitoring of users' activities to confirm their identities using particular user profiles that are modeled using AI techniques. To carry out the experiment, a unique application was created to collect user data in a supervised situation in which certain tasks must be finished in advance. After anonymization, this dataset will be made accessible to the public. Furthermore, a publicly available dataset was utilized for benchmarking, enabling our methods to be verified in an unguided environment. These data were processed to identify several important properties that might be utilized for training three distinct AI methods: Multi-Layer Perceptrons, Support Vector Machines, and a deep learning network. These methods proved to be successful in both situations and were able to effectively authenticate users. To detect imposters when an authenticated session is hijacked in a real-world setting, a continuous authentication method was designed and tested utilizing weighted sliding windows, and a rejection test was finally carried out.

Keywords: Supervised, AI, Deep Learning, Authenticated, Windows,

1. INTRODUCTION:

Since the beginning of information technologies (IT), one of the primary problems and research areas has been computer security, or cybersecurity. The term "cybersecurity" describes the defences put in place for an IT system to maintain the availability, integrity, and confidentiality of its resources [1, 2]. Numerous security protocols have been created over time to safeguard various kinds of assets, including data and software services as well as tangible items. A lot of these approaches need the integration of real-time data processing in a Big Data context with Artificial Intelligence (AI) capabilities. We have presented a system for intrusion detection (IDS) based on machine learning approaches, and our team of researchers has proven its skill in this sector [3, 4]. We employed self-organizing maps (unsupervised neural networks) to identify and separate potentially dangerous aberrant network events. We have presented a system for intrusion detection (IDS) based on machine learning approaches, and our team of researchers has proven its skill in this sector [3, 4]. We employed self-organizing maps (unsupervised neural networks) to identify and separate potentially dangerous aberrant network events. Despite these advancements, there is still no mechanism to identify identity theft during an open session once the user has been given access and privileges through these techniques. Periodically requesting authentication from the user could help to mitigate this problem. But using a shared approach for this would be impractical because it would keep interrupting users' sessions. Using user behaviour as a second-phase verification focused on the ongoing monitoring of the user-system interaction is an additional or alternative method to achieve this [13]. In this regard, the easiest devices to evaluate first are standard input/output devices like keyboards and pointing devices [8]. If the continuous monitoring identifies an abnormality in the users' behaviour, it may prompt a second password request or the delivery of an SMS code to confirm the users' identification. It may also send an alarm to the system's administrators.

2. LITERATURE SURVEY

F. Chong et. al [23] proposed that an identity is described as comprising of qualities, attributes, and preferences upon which one may obtain tailored services which might exist online, on mobile devices, at work, or in many other locations. It goes on to categorise the many types of identification traits into three basic groups: biometrics, physical metrics, and pseudo metrics. Biometrics is the automated technique for measuring and assessing an individual's physical and behavioural traits. Fingerprint, facial, and iris scans are among examples. Physical metrics refer to what we have, which includes all physical credential tokens such as personal computers, mobile phones, and card-based credential tokens such as smart cards. In this article, the physical tokens' identifying properties shall be referred to as

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8
Article Received: *20 May 2023* ***Revised:*** *28 May 2023* ***Accepted:*** *16 August 2023*

device metrics. The device metrics include the IP address, International Mobile Equipment Identifier (IMEI), Subscriber Identify Module (SIM), and a unique card identification number. The pseudo metrics encompass all identify traits that fall under the heading of "something you know." Password and personal identification number (PIN) are excellent examples. The three categories are utilised in this research to create a multifactor authentication system using information fusion, in which the user must input an identifying attribute from at least one of the three groupings. Identity theft and identity fraud are phrases used to describe various forms of crimes in which someone illegally gets and utilises another person's personal data for financial advantage. Cybercrime has emerged as one of the world's most rapidly expanding crimes. Identity fraud has become a big problem for both the public and corporate sectors, especially in relation to terrorism, money laundering, financial crime, drug trafficking, immigrant smuggling, and arms smuggling. To solve these security concerns, M. Hansen et al. [27] created virtual identities based on service sessions to safeguard the user's privacy from both service providers and access network providers. Another milestone is frameworks that allow users to track the history of how their identification information is treated once it is transmitted between domains of control. G. Hidehito et al. [29] propose a method for privacy-controlled exchange of identification characteristics in a federated setting. This research introduces statistical and artificial intelligence approaches in a multifactor authentication system to aid in the prevention of cybercrime.

3. OBJECTIVES:

A continuous authentication technique needs a strong feature set that enables precise identity verification of users almost instantly, in addition to a non-intrusive method of monitoring users. Therefore, these factors must be considered to construct a workable authentication method; otherwise, it will not be appropriate for use in real-world settings. The primary objectives that we purport to achieve in this study can be summed up as follows:

- Create an entirely transparent data collection process for the user to ensure that it does not impede their activity. As a result, data collection needs to be done efficiently so as not to interfere with user experience or add undue computing burden to the system. This implies that data collecting needs to be sufficiently light to be sent across the communications network effectively.
- Determine a collection of pertinent or essential characteristics that can be utilized to conclusively confirm users' identities based on how they behave when using a shared pointer device. The real owner of the active user account is the identity that needs to be validated, as our aim is to create a second-phase authentication system.
- Therefore, to identify any session hijacking, the activity that is being monitored during a given session can be compared to a pre-built profile of the user for the actual account owner. To avoid identity usurpation, the entire process—from data collection to identity verification—must be completed in close to real-time. As a result, even though the chosen techniques are computationally possible, they should not jeopardize the authentication process' correctness. Specifically, the process of extracting features and verifying identity using pre-made user profiles are crucial elements in terms of efficiency.

4. PROPOSED FLOWCHART

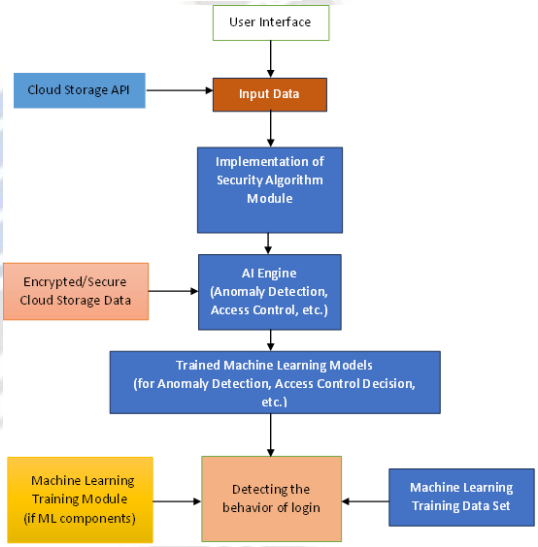

Fig. 1.1 Flow of the proposed system

A. Cloud Storage API

This API is designed and implemented to stored the data in authenticate cloud when the behavioural patterns match with authenticate user and if then pattern of authentication model doesn't matches with the user, then it will not allow to store any data into the cloud. An API library was used to integrate the data with cloud.

B. Security Algorithm modules

A specific algorithm was designed to detect the behaviour of the user while authenticating into the account and this behaviour patterns allows to create the sequence of authentication before login into the account.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8
***Article Received:** 20 May 2023 **Revised:** 28 May 2023 **Accepted:** 16 August 2023*

C. Anomaly Detection

A pattern of anomaly was already available in database to predict the change in patterns while using the account and in case it helps to compare the unreal behaviour with methos of the user authentication.

D. Trained Models

The models were trained and created by checking the patterns of user input while authenticating and uses the same pattern for training and testing using the algorithm for creating the model for further comparison of the input sequence.

E. Behaviour Detection

The proposed final result predicts the user authenticity depends upon its sequence of authentication procedure and helps to avoid various attacks that was generated during any login process.

5. FUTURE OF AI AND AUTHENTICATION

There will be further advancements and sophistication in risk-based authentication. AI-powered authentication will probably eventually transition from supervised learning—where the dataset contains the outcomes—to unsupervised learning, where the AI looks for new patterns to utilize in order to provide predictions that it may not have found in the human world. The accuracy and range of AI-based authentication services can be enhanced by cross-referencing several machines learning algorithms, applying pattern recognition, and utilizing time-series-based prediction algorithms. Simultaneously, developers will want to provide IT departments additional oversight over the AI system. This control might include the ability to ascertain precisely which data was used in a particular decision, modify the number of criteria considered, and customize the system to the specific environment of their firm. One area that businesses, such as OneLogin, are currently looking at is the capacity to use data from other parties. A stolen credential check that leverages third-party data on compromised or exposed credentials is part of OneLogin's Smart Factor Authentication. Furthermore, several cross-industry initiatives are in motion to facilitate improved data-sharing, allowing one organization's knowledge of a possible threat to be instantly shared with other businesses.

6. CONTINUOUS AUTHENTICATION SCHEME

It is important to emphasize that none of the above-mentioned models can function independently of the environment that manages the entire process. This environment gathers and processes mouse data to extract features, integrates the models, and creates a continuous authentication system that detects potential intrusions based on predictions made for each chunk.

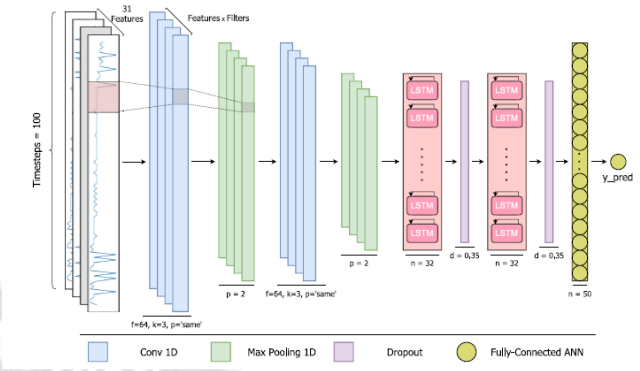

Fig. 1,1 Deep learning architecture overview

First, it was discovered that the fixed-size strategy outperformed time-based chunks by a little margin. This is because there is a lot of variation in the number of records that are available every chunk. For example, during a loading window wait period, the user may travel for several seconds without moving at all. Other times, the user may move frequently in a brief amount of time. As a result, it appears that time-based chunking is inappropriate in these circumstances; instead, fixed-size chunks have shown to be more reliable and effective. We choose to concentrate on the newest kind of chunks (fixed-size) as a result. We discovered that their performance was comparable across the various sizes that were considered (100, 200, 500). However, using small-sized pieces will enable us to carry out the authentication process more frequently than using bigger portions, while lowering the computing cost of the entire process, as our goal is to identify users continuously throughout their entire user sessions. As a result, we are reporting the findings for both datasets in fixed-size pieces of one hundred incidents for each of the suggested approaches. Reducing false positives is the primary goal of authentication; that is, to identify any system intrusion, which may be effectively assessed using the accuracy score or false positive rate (FPR). But it is also important to consider false negatives, as continually rejecting a valid user can create annoyance and disturb their normal activity.

7. CONCLUSION

Following the proper configuration, the models' performance was assessed using a variety of commonly used metrics, including recall, precision, and the F1 score. To begin with, a cross-validation process was run to get the baseline findings that show the suggested approaches can authenticate users that are legitimate. Subsequently, a denial test was conducted on the authentic user models constructed in the preceding phase to assess their potential to identify and avert outsider

user session takeover. In this work, a comprehensive experiment was carried out to evaluate the capacity of behavioural of login procedure attributes to verify the user's identity using AI approaches. To conduct the experiment in a directed scenario, we first constructed and distributed a non-intrusive login model program participant. In this scenario, each user was given a series of tests including various mouse operations. Then, using three distinct AI techniques—MLP, SVM, and DL—a vast range of mouse movement data were retrieved, examined, and analysed to create unique user profiles based on the data that was acquired.

REFERENCES

[1] S. Mike, Unify and Simplify: Re-thinking Identity Management, Network Security. 2006(7) (2006), 11-14. doi: dx.doi.org/10.1016/S1353-4858(06)70411-1

[2] R. Dhamija, and L. Dusseault, The Seven Flaws of Identity Management: Usability and Security Challenges, Security & Privacy, IEEE. 6(2) (2008) 24-29. doi: dx.doi.org/10.1109/MSP.2008.49.

[3] S. Clare, Digital identity - The Legal Person? Computer Law & Security Review, Elsevier. 25(3) (2009) 227-236. doi: dx.doi.org/10.1016/j.clsr.2009.03.009.

[4] P. Geraint, The benefits and drawbacks of using electronic identities, Information Security Technical Report, Elsevier. 13(2) (2008) 95-103. doi: dx.doi.org/10.1016/j.istr.2008.07.002.

[5] G. Goth, Identity management, access specs are rolling along, Internet Computing, IEEE. l9(1) (2005), 9- 11. doi; dx.doi.org/10.1109/MIC.2005.16.

[6] B. Geoff, The use of hardware tokens for identity management, Information Security Technical Report 9(1) (2004) 22-25. doi: dx.doi.org/10.1016/S1363- 4127(04)00012-3.

[7] H. Marit, P. Andreas and S. Sandra, Identity management throughout one's whole life, Information Security Technical Report, Elsevier. 13(2) (2008) 83-94. doi: dx.doi.org/10.1016/j.istr.2008.06.003.

[8] EconomyWatch, List of Commercial Banks; Available (January 2011)

[9] Wikipedia, The free encyclopedia, Civil service; Available (January 2011): http://en.wikipedia.org/wiki/Civil_service

[10] Top University, QS World University Rankings; Available (January 2011): http://www.topuniversities.com/universityrankings/world-university-rankings, 2011

[11] Thomson Reuters, 100 Top Hospitals; Available (January 2011): http://www.100tophospitals.com/

[12] Google Double Click Ad Planner; Available (April 2010): http://www.google.com/adplanner/static/top1000/#

[13] Wikipedia, The free encyclopedia, List of social networking websites; Available (January 2011): http://en.wikipedia.org/wiki/List_of_social_networking

[14] L. Anthony. AntConc: Design and Development of a Freeware Corpus Analysis Toolkit for the Technical Writing Classroom, in Professional Communication Conference Proceedings, (2005), pp. 729. doi:10.1109/IPCC.2005.1494244

[15] C. Greaves, ConcApp Version 4 Concordancer, Edict Virtual Language Centre, Available (November 2010): http://www.edict.com.hk/PUB/concapp/.

[16] TextSTAT Corpus, Available (November 2010) on: http://www.edict.com.hk/PUB/concapp/ Published by Atlantis Press Copyright: the authors 429

[17] Neural-fuzzy multifactor authentication system 17. R. Togneri, and C. J. S. DeSilva, Fundamentals of Information Theory and Coding Design, (Chapman & Hall/ CRC Press, FL, 2005).

[18] M. Negnevitsky, Artificial Intelligence: A Guide to Intelligent Systems, 2nd edn. (China Machine Press, 2005)

[19] M. Fazle Azeem, M. Hanmandlu, N. Ahmad, Structure identification of generalized adaptive neuro-fuzzy inference systems, Fuzzy Systems, IEEE Transactions. 11(5) (3003) 666–681. doi:10.1109/TFUZZ.2003.817857.

[20] J. I. Agbinya, R. Islam and C. Kwok, Development of Digital Environment Identity (DEITY) System for Online Access, in Broadband Communications, Information Technology & Biomedical Applications, Third Int. Conf., (Australia 2008), pp. 23-26, doi: 10.1109/BROADCOM.2008.52.

[21] M. He, et al, Performance Evaluation of Score Level Fusion in Multimodal Biometric Systems, Pattern

[22] Recognition. 43(5) (2010) 1789-1800. doi: 10.1016/j.patcog.2009.11.018.

[23] L. Nanni, A. Lumini, S. Brahnam, Likelihood Ratio Based Features for a Trained Biometric Score Fusion, Expert Systems with Applications. 38(1), (2011) 58-63. doi: 10.1016/j.eswa.2010.06.006.

[24] F. Chong, Identity and Access Management, Microsoft Architect Journey. (2004).

[25] European Technology Assessment Group, RFID and identity management in everyday life, Available

online (October 2010) at: http://www.europarl.europa.eu/stoa/publications/studies/s toa182_en.pdf

[26] The National Electronic Commerce Coordinating Council (NECCC), Identity Management, Presented at the NECCC Annual Conference, (New York, 2002).

[27] M. Hansen, A. Schwartz and A. Cooper, Privacy and Identity Management, Security & Privacy, IEEE. 6 (2008). doi: 10.1109/MSP.2008.41.

[28] M. Barisch, Modelling the Impact of Virtual Identities on Communication Infrastructures, Conference on Computer and Communications Security, in Proc. of the 5th ACM workshop, Digital identity management, (Chicago, Illinois, 2009) pp. 45–52. doi: http://doi.acm.org/10.1145/1655028.1655040.

[29] G. Hidehito, User-Centric Identity Governance Across Domain Boundaries, Conference on Computer and Communications Security, in Proc. of the 5th ACM workshop, Digital identity management, (Chicago, Illinois, 2009) pp. 35–44. doi: http://doi.acm.org/10.1145/1655028.1655038.

[30] V. Avram, Defining metrics to automate the quantitative analysis of textual information within a web page, in Int. Conf. of Application of Information and Communication Technologies, (AICT 2009), pp.1-5. d