World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266

Cross Ref DOI: 10.30574/wjaets

Journal homepage: https://wjaets.com/

(Review Article)

A study of risk-based authentication system in cyber security using machine learning

Imran M. Hussain Qureshi * and Vijay K. Kale

Dr. G Y Pathrikar College of CS and IT, MGM University, Aurangabad, Maharashtra, India.

World Journal of Advanced Engineering Technology and Sciences, 2022, 07(02), 065–070

Publication history: Received on 28 September 2022; revised on 05 November 2022; accepted on 08 November 2022

Article DOI: https://doi.org/10.30574/wjaets.2022.7.2.0125

Abstract

The optimum authentication method is determined by the user's risk profile, which is created using context- and behavior-based data from the user's device, finger print, one-time password, and other characteristics. Hacking and security breaches of online accounts, including social networking and web ac- counts, are very common in today's society. We suggest a Risk Based Authentication System utilizing Machine Learning to stop this. For the protection of data and money in this internet environment, security is a worry. Numerous parameters are researched and taken into consideration in the paper in order to solve the issue. These variables determine whether to grant the user permission or not. The gradients descent method is used to verify the user. Previous literature is re- viewed with technical details of the system before conclusion.

Keywords: Authentication; Risk; Machine Learning; Gradients Descent; Security

1. Introduction

Adaptive authentication is a technique engaged to choose the authentication step based on the risk factor identified by the system during the first phase of login procedure. A one-time password or code verification through a smartphone will be activated if the risk appears to be reasonable. As long as nothing out of the ordinary occurs, like a login attempt coming from a brand-new machine, the user will benefit from this [1]. Despite being widely used, passwords are insecure, have faults, and users do not like them. This is an area of authentication research that has been highly active and demonstrating for many years. Numerous studies have been conducted to develop alternatives to passwords and to lessen the difficulty of managing them. One crucial component of information system security is authentication. Protecting resources from access by un- authorized users is the first stage in an access control method. A system identifies a user by asking for their credentials, which are subsequently verified against some command. Users who can successfully authenticate themselves are the only ones to whom authorization is granted [2].

According to their credentials, the various types of authentication processes can be categorized into the following groups: holding on to something handed to the user; sig- nature; mouse; and fingerprint [3]. Based on the user's risk profile, which is created using information about their behavior and their context, adaptive authentication is a trustworthy technique for dynamically choosing the best techniques among different modalities to authenticate a user. Using adhoc or adaptive authentication methods, rule- based techniques, and user behavior and communication patterns, existing systems evaluate user activity. These solutions are operationally useless and no one model is adequate for a global attack due to the dynamic nature of internet fraud strategies [4]. A necessary precaution against unauthorized access to the systems, devices, and other means of sensitive applications is authentication or testimony. The single-factor authentication, often known as the testimony, is unreliable and easily circumvented. Adaptive authentication, sometimes known as testimony, enables a system to dynamically choose the appropriate authentication

[*] Corresponding author: Imran Qureshi
Dr. G Y Pathrikar College of CS and IT, MGM University, Aurangabad, Maharashtra, India.

methods according on the user's context, including their signature, fingerprint, device accessibility, and other characteristics.

This strategy is knowledge-based, making it easily prone to dangers. This static, single-factor authentication does not, therefore, provide 100 percent security. Addition- ally, the gadget makes no calculations to determine whether the user is legitimate or not. On the other hand, implementing biometrics is expensive. Using login parameters and credentials together can increase security [5]. Eight user parameters that are gathered during the current login attempt for a particular user are used to examine a user login record. Before utilizing login credentials and other login attributes, a user is analyzed based on his prior habits and categorized as legitimate or fake. These characteristics are examined using a machine learning model with three training modules that calculates risk level and selects the appropriate authentication methods for a given user. Utilizing three algorithms improves classification accuracy. As a result, a real user won't have to go through many randomly generated sets of authentication mechanisms. This guarantees that the system's security and usability are upheld. Since the user can- not guess the authentication method, this also gives the authentication system a dynamic quality. A low-cost, low-computational authentication method is promised by this idea [6].

2. Literature Review

The many techniques that have been utilized to study user behavior that falls under the heading of adaptive security are explained in this part. The system is also examined in order to update the security defense at run time. The authentication research field is quite active and has been proving for more than 42 years that, despite being widely used, passwords are inadequate, unsafe, and rejected by users. One of the fundamental problems with information system security is authentication or testimony. It is a crucial part of the access control process that guards against unauthorized users accessing re- sources. A crucial defense against unauthorized access to systems, devices, and additional critical applications is authentication or testimony. The adaptive systems com- munity is a wonderful place to start when looking for engineering tools as well as adaptive authentication or testimony designs. This tool offers assistance in creating the soft- ware for adaptive systems and can serve as the basis for the definition of similar tools that take into account the specifics of the authentication or testament domain [7].

To simulate user behavior, a device is employed in a spatial-temporal environment. The explicit, implicit, and session profiles were used to study user behavior. Users are di- vided into three categories: normal, suspicious, and abnormal. A Bayesian model for user mouse dynamics was created. The single-factor authentication is not dependable and is relatively simple to get around. Although there have been recent in-depth investigations of password substitutes, no work has yet given a comprehensive assessment of the combined usage of several authentication or testimony procedures inside adaptive systems [8]. In the set of traits chosen for recognizing a person's fingerprint based on the context, there are variables that are used for comparison in adjusting behavior. Online applications and systems are progressively being exposed to several cyber dangers. This is mostly result of these systems' widespread use of traditional authentication techniques like login and password. These systems are particularly vulnerable to as- saults like phishing, DNS, and mask attacks since they have access to many hacking tools. Additionally, passwords are typically knowledge-based information that people can exchange. Therefore, validity of users cannot be totally guaranteed by single factor authentication. Although users may find single factor authentication to be simple, two factor authentication, multifactor authentication or testament are preferable owing to their higher security standards [9].

A design for dynamically choosing elements for multifactor authentication was created in a recent study. For each factor, accurate values were determined using mathematical objective functions. The choice of elements was influenced by the media, the device, and environmental circumstances including light and noise. This method reduced the need to repeatedly choose the same group of authentication factors [10]. The construction of systems with controlled resources that are devices and applications with heterogeneous authenticators can be approached by translating the concepts from the classification of self-adaptive systems design to adaptive authentication. However, the authentication domain has never seen the application of this kind of systematic design approach. The development of authentication or testimony systems that respond to environmental changes creates new attack vectors and points of vulnerability. Potential security hazards were identified in half of the publications surveyed [11]. Device theft and secure pairing when using tokens as authenticators are the difficulties that have been identified. Data privacy, particularly if it involves data that is not stored locally or calculations that are outsourced to a third party. Attacks made against machine learing based system components, such as those that force an unauthorized user to be mistakenly classified as authentic.

Techniques for securely outsourcing continuous implicit authentication systems are sought in order to overcome the privacy concerns that arise from the implicit mechanisms situation. Additional difficulties brought on by continuous authentication are linked to the use of machine learning by the device [12]. All of these decisions influence authentication performance generally and should be improved. Since adaptive authentication systems will still require password manipulation, it would be exciting to investigate their connection with existing APIs that allow for programmatic access to user credentials maintained in password managers to speed up the sign-in process. User role, Authentication preferences, and Modifications on users are crucial components to take into account when it comes to user-related changes [13].

3. System Design

Three parts make up the dynamic authentication system i.e. creation of data, risk engine, and authentication. Following procedures are included in the data generation module to prepare applicable dataset from raw dataset.

Account creation: It requires the user to enter accurate information such as an email address, a mobile number, a pattern lock code, the answers to security questions, and a graphical password. This is the first step in the registration process that must be completed in order to accept the credentials that will subsequently be used to validate the user's entrance.

Extraction of parameters: A server-side validation system verifies, upon user input of the username, the existence of the user account. If so, the session is used to extract real- time login parameters including OS, Browser, IP, Device, Time Zone, Login Time, Geo Location, and Number of Failed Attempts.

Data preprocessing: Records' input contains the attribute Login Time, which is time zone dependent. We add a third argument called Global Standard Time, which trans- forms the login time into standard format in order to stop this dependency between these two parameters (G.S. T). When doing additional calculations, login time is taken into account in accordance with global standard time once the time zone and login time have been eliminated. Since all ML algorithms require numeric data as input, encode the data.

User History Retrieval: Past login information for a specific user has been retrieved from a database according to username for additional analysis. For additional analysis and the computation of risk ratings, the Risk Engine module receives the fetched history and extracted parameters from the current login.

Support Vector Machine (SVM): Supervised Machine Learning can be applied to classification and regression issues. SVM is a labeled data-based supervised machine learning algorithm. SVMs identify a hyper plane that efficiently separates a dataset into several classes. [10]

One Class SVM: As its name suggests, one-class SVM classification is a unary classification or class-modeling [11] that tries to identify items belonging to a given class among all objects inputted by learning from a training set including only the objects of that class. The training set contains items from all the classes, making it harder than previous approaches to distinguish between two or more classes. One-class SVM only considers one class, the novelty/normal class.

4. Risk-Based User Authentication

Risk authentication systems, as shown in Fig. 1, are based on a continuous choice to approve or reject user authentication by observing the user's behavior and the risk of his or her activity, according to review article [12]. The system then prompts the users for re-authentication based on the comparison of a risk score computed in real time with the stored risk profiles of the users. Re-authentication should not be necessary, for ex- ample, when an officer is using the mobile identifying device from a confirmed secure location (a workplace for land or sea border control). While the service may need more proof of the user's identity and hence request re-authentication in the case of an unidentified or unverified location. Risk-based authentication techniques are becoming more popular since they tend to provide frictionless user authentication while boosting security and encouraging user comfort [12]. Although there are numerous security businesses that offer risk-based authentication for mobile devices, the technology that the risk-based authentication scheme employs to compute the risk score—the risk estimate scheme—makes the difference in whether a mechanism is effective or not [13]. In order to create a risk score that is accurate and reliable while causing the least amount of disruption to the user's experience, an effective risk-based authentication solution will not only use contextual user information (such as the device's ID, location, date, time, and connection) but will also use the user's behavioral patterns [14], the device attributes, the user history, and other factors.

Additionally, it is crucial to use risk estimate techniques in order to accurately calculate the risk score. Although existing qualitative techniques appear to be plausible, they rely heavily on expert intuition and always rank hazards subjectively, rendering them inappropriate for use in sensitive applications like public safety and real-world scenarios [15]. Finally, proactive or reactive approaches can be used to implement risk-based user authentication [12]. In the first scenario, a risk-based authentication scheme actively foresees the emergence of potential attacks, failures, or any other type of security vulnerabilities and promptly takes action. Re-authentication is necessary when the risk score rises above the permissible threshold level, whereas reactive risk-based authentication accepts some risks up until that point.

Additionally, it is crucial to use risk estimate techniques in order to accurately calculate the risk score. Although existing qualitative techniques appear to be plausible, they rely heavily on expert intuition and always rank hazards subjectively, rendering them inappropriate for use in sensitive applications like public safety and real-world scenarios [15]. Finally, proactive or reactive approaches can be used to implement risk-based user authentication [12]. In the first scenario, a risk-based authentication scheme actively foresees the emergence of potential attacks, failures, or any other type of security vulnerabilities and promptly takes action. Re-authentication is necessary when the risk score rises above the permissible threshold level, whereas reactive risk-based authentication accepts some risks up until that point.

Please try to avoid rasterized images for line-art diagrams and schemas. Whenever possible, use vector graphics instead (see Error! Reference source not found.).

5. Machine Learning Classification Algorithms

A decision tree is a supervised classification technique that uses a set of rules provided in a tree-like structure to execute hierarchical decision making on the feature values [16]. Data are first split into training and validation data sets, as is the case with all machine learning methods. Using methods like recursive partitioning, training data are utilized to determine the right set of rules and the best division for particular qualities. The training data is divided into two or more branches by any choice that splits the tree based on a criterion. Finding the best split criterion is the key goal in order to minimize the number of class variables that are mixed in each branch of the tree [17].

The decision tree is then validated, and any modifications required to make the tree more effective are made using the validation data [16]. The three traditional decision tree implementation algorithms are ID3, C4.5, and CART (Classification and Regression Trees). These algorithms use the splitting criterion "entropy." To develop their data model, they used the CART classification algorithm among the three aforementioned traditional decision tree algorithms because the output of their system is a binary decision: Accept or Deny authentication. Additionally, they created a dataset using Matlab based on cutting-edge research to create their data-driven model (i.e., classifier). In terms of reducing over-fitting and being able to handle missing data, CART outperforms other algorithms [17].

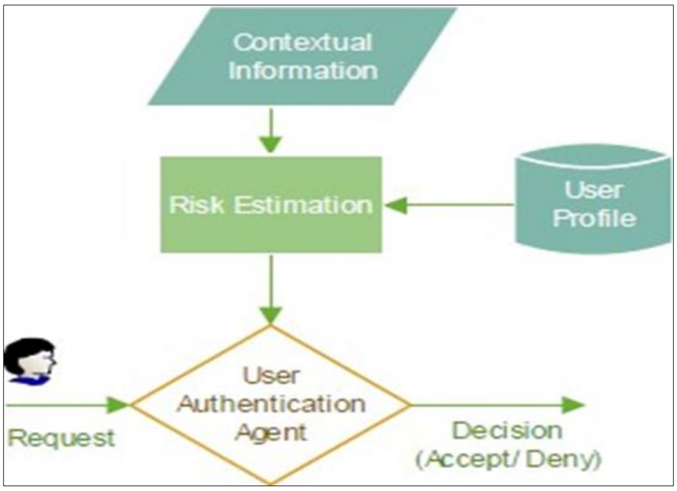

Figure 1 Risk-based user authentication overview [2]

Additionally, it has the ability to create models for both classification and regression. The Gini criterion is used by CART to divide the training data. Scikit Learn's implementation of CART has been optimized, as seen in [17]. The fundamental benefit of decision trees is that, provided a suitable set of rules is established by security professionals, they function well even in the absence of sufficient data. However, the division of feature values in a standard decision tree model is based on classical set theory. Furthermore, due to the partition's discreteness, a small change in a particular attribute's value could result in a completely different conclusion. Decision trees are regarded as useful and simple categorization models [16]. However, as the decision tree grows in size, comprehension of it becomes more challenging, and more data are also required for identifying and evaluating the set of rules [18]

6. Conclusion

Before selecting whether or not to provide access to the user, a variety of different factors have been taken into account. We will verify that if the gradients are close to zero, the user will not be granted access to the system; however, if the gradients are close to one, the user will be granted access to the system. Here, we must examine thedesign principles used in the field of adaptive authentication systems to study the literature on adaptive authentication. We anticipate that our study will serve as a starting point for understanding adaptive authentication systems and advancing related re- search.

Compliance with ethical standards

Acknowledgments

We certify that the article is the Authors' original work. The article has not received prior publication and is not under consideration for publication elsewhere. On behalf of all Co-Authors, the corresponding Author shall bear full responsibility for the submission. This research has not been submitted for publication nor has it been published in whole or in part elsewhere. We attest to the fact that all Authors listed on the title page have contributed significantly to the work, have read the manuscript, attest to the validity and legitimacy of the data and its interpretation, and agree to its submission to World Journal of Advanced Engineering Technology and Sciences (WJAETS).

Disclosure of conflict of interest

The author(s) certify that they have No Conflict of Interest in the subject matter or materials discussed in this manuscript.

References

[1] Anne Adams and Martina Angela Sasse (1999) Users are not the enemy. Commun. ACM 42, Vol. 12, pp-40–46.

[2] Aditi Gupta, Markus Miettinen, N. Asokan and Marcin Nagy (2012) Intuitive security pol- icy configuration in mobile devices using context profiling, in Proceedings of the Interna- tional Conference on Privacy, Security, Risk and Trust (PASSAT'12) and the International Conference on Social Computing (SocialCom'12) IEEE, pp-471–480.

[3] Christian Krupitzer, Felix Maximilian Roth, Sebastian Van Syckel, Gregor Schiele and Christian Becker (2015) A survey on engineering approaches for self-adaptive systems, Pervas Mobile Computation, Vol. 17, pp-184–206.

[4] Alain Forget, Sonia Chiasson and Robert Biddle (2015) Choose your own authentication, in Proceedings of the New Security Paradigms Workshop, pp-1–15.

[5] Arun Ramakrishnan, Jochen Tombal, Davy Preuveneers and Yolande Berbers (2015) PRISM: Policy-driven risk-based implicit locking for improving the security of mobile end-user devices, in Proceedings of the ACM International Conference on Advances in Mobile Computing & Multimedia (MoMM'15), pp-365–374.

[6] Abdeljebar Mansour, Mohamed Sadik, Essaid Sabir and Mohamed Azmi (2016) A context- aware multimodal bio-metric authentication for cloud-empowered systems, in proceedings of the IEEE International Conference on Wireless Networks and Mobile Communications (WINCOM'16). 278–285.

[7] Adam Wojtowicz and Krzysztof Joachimiak (2016) Model for adaptable context-based bi- ometric authentication for mobile devices, Pearson Ubiq Computing Vol. 20, Issue 2, pp- 195–207.

[8] Adam Wojtowicz and Jacek Chmielewski (2017) Technical feasibility of context-aware passive payment authorization for physical points of sale, Pearson Ubiq Computing, Vol. 21, Issue 6, pp-1113–1125.

[9] Blase Ur, Felicia Alfieri, MaungAung, Lujo Bauer, Nicolas Christin, Jessica Colnago, Lorrie Faith Cranor, Henry Dixon, Pardis Emami Naeini, Hana Habib et al. (2017) Design and evaluation of a data-driven password meter, in Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'17), pp-3775–3786.

[10] G. Mantas, N. Komninos, J. Rodriguez, E. Logota, and H. Marques, Security for 5G Com- munications, in Fundamentals of 5G Mobile Networks, J. Rodriguez, L. Eds., John Wiley & Sons, Ed. Chichester, UK, 2015, pp. 207–220.

[11] M. Papaioannou et al., A Survey on Security Threats and Countermeasures in Internet of Medical Things (IoMT), Transmission Emerging Telecommunication Technology, no. May, pp. 1–15, 2020.

[12] S. Gupta, A. Buriro, and B. Crispo, Demystifying Authentication Concepts in Smartphones: Ways and Types to Secure Access, Hindawi Mob. Inf. Syst., vol. 2018, 2018.

[13] A. J. Harris and D. C. Yen, Biometric authentication: Assuring access to information, Information Management Computer Security, vol. 10, no. 1, pp. 12–19, 2002.

[14] J. Ribeiro, F. B. Saghezchi, G. Mantas, J. Rodriguez, and R. A. Abd-Alhameed, HIDROID: Prototyping a Behavioral Host based Intrusion Detection and Prevention System for An- droid, IEEE Access, vol. 8, pp. 23154 – 23168, 2020.

[15] J. Ribeiro, F. B. Saghezchi, G. Mantas, J. Rodriguez, S. J. Shepherd, and R. A. Abd- Alhameed, Towards an Autonomous Host-based Intrusion Detection System for Android Mobile Devices, in 9th EAI International Conference on Broadband Communications, Net- works, and Systems (BROADNETS2018), 2018, pp. 139–148.

[16] H. F. Atlam, A. Alenezi, R. J. Walters, and G. B. Wills, An overview of risk estimation techniques in risk-based access control for the internet of things, IoTBDS 2017 - Proc. 2nd Int. Conf. Internet Things, Big Data Security, no. April, pp. 254–260, 2017.

[17] M. Heydari, A. Mylonas, V. Katos, E. Balaguer-Ballester, V. H. F. Tafreshi, and E. Benkhelifa, Uncertainty-aware authentication model for fog computing in IoT, in Fourth International Conference on Fog and Mobile Edge Computing (FMEC), 2019, pp. 52–59.

[18] S. Wang, C. Fan, C.-H. Hsu, Q. Sun, and Y. Fangchun, A Vertical Hand off Method via Self- Selection Decision Tree for Internet of Vehicles, IEEE Syst. Journal, 10(3), pp. 1183– 1192, 2016.