International Journal of Soft Computing and Engineering (IJSCE)
ISSN: 2231-2307 (Online), Volume-10 Issue-6, July 2021

A Systematic Approach for a Secure Authentication System

H A Gautham, Ramakanth Kumar P

***Abstract*: Authentication is a process of verifying the credibility of a user who is trying to access classified or confidential information. There is a vast unfold in the number of internet users, and the demand for IoT devices, cloud services has been increasing; it is now essential more than ever to protect the data hosted on the internet. So, the authentication process cannot be relied on single-factor static authentication methods to verify the user credentials. All devices in the market are not equipped with biometric systems, so a form of multi-factor authentication which is independent of biometrics needs to be adopted for a secure authentication system. This paper portraits a systematic architecture to verify user credentials using specific parameters, trying to unfold patterns using machine learning algorithms based on user's past login records, thus trying to provide a safer and secure authentication process for the users.**

***Keywords*: Authentication Process, Machine Learning, Multi-Factor Authentication, Security, Two-Factor Authentication, User Login.**

I. INTRODUCTION

Plenty of cyber-attacks are currently being conducted, and it is only increasing over time. Most of the current password-based authentication methods expose weakness, making most systems susceptible to various attacks such as DNS attacks and phishing attacks. Still, password-based authentication schemes are one of the presiding authentication mechanisms which are used by online systems today. Rather than creating new authentication mechanisms, it is more critical to enhance the current password-based mechanisms. To enhance the current mechanisms, various additional factors are considered to verify the user's credibility. Such authentication mechanisms are called multi-factor authentication (MFA).

Multi-factor authentication schemes can be of different forms, such as expecting another credential like biometric identity or a pin, or it can even be using certain user's current attributes to evaluate the user's risk profile before granting access to the user. The latter is popularly known as Risk-based authentication (RBA). It might seem that a single-factor authentication process is faster and easier for the user, but it comes with a trade of decreased security levels.

Fig. 1. Multi-Factor Authentication.

RBA is a form of multi-factor authentication designed to protect users by analyzing specific user attributes. RBA plays a vital role in the systematic approach for a secure authentication system presented in this paper. Machine learning algorithms are used to analyze the user's profile, where it classifies whether the user is fraudulent or not with a specific probability. That probability measure is then used to challenge the user further to verify the user's credibility if required. Thus, having a perfect balance between security and ease of use with the additional benefit of easy upgradability for the current password-based systems.

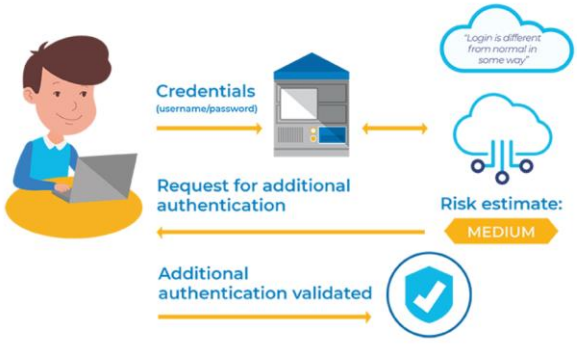

Fig. 2. Risk-based Authentication.

Manuscript received on June 06, 2021.
Revised Manuscript received on June 11, 2021.
Manuscript published on July 30, 2021.
* Correspondence Author
H A Gautham*, Department of Computer Science and Engineering, R.V. College of Engineering, Bengaluru (Karnataka), India. Email: gautham1327@gmail.com
Dr. Ramakanth Kumar P, Professor & Head, Department, Department of Computer Science and Engineering, R.V. College of Engineering, Bengaluru (Karnataka), India. Email: ramakanthkp@rvce.edu.in

II. LITERATURE SURVEY

This section briefs and analyses current research work related to the topic. A study is presented in the paper by Wiefling et al [1] on eight popular online services analyzing which features and classifiers are used in RBA. A framework is also presented in the above paper to measure the performance of RBA and concluding with throwing light on how RBA is becoming more and more important to strengthen password-based authentication. A mathematical and statistical supported approach to strengthen password-based authentication by finding out suspicious activity based on certain parameters is presented in the paper by Freeman et al. [2]. A light is thrown on how password reusability puts multiple accounts used by the user at risk in a paper presented by Bailey et al [3]. It has also been stressed in the above paper that users re-use passwords even for high valued accounts. Weak passwords are a predominant problem, and their vulnerability against guessing attacks has been shown in a paper presented by Bishop et al [4]. A multi-factor evaluation process based on weight given to certain user's parameters is presented in the paper by Diep et al [5]. A fuzzy logic-based approach is presented to evaluate the risk profile of the user is presented in the paper by Cheng et al [7]. In the existing systems, there is a lack of use of modern machine learning algorithms, which have become increasingly popular due to their high accuracies. These differences shall be addressed by the approach presented in this paper.

III. DETAILED ARCHITECTURE

This section covers the detailed architecture which should be followed systematically to attain a secure authentication system. The architecture has two parts: Fraudulent classifier and Risk analyzer. The following sub-sections cover both parts.

A. Fraudulent Classifier

The user sends an authentication request with his username-password credentials embedded in HTTP headers or a custom login process. Along with that following user, parameters are retrieved from the request and previous records.

- IP Address
- Availability of cookie
- Browser and its version
- OS and its version
- Unsuccessful attempts of login
- IP location (Available from IP2Location [6])
- Time zone or country (whichever is available)
- Accept-language
- Display resolution
- Login time

Each login from the user serves as an input for the next login and improves the efficiency of the fraudulent classifier. So, every login attempt by a user is stored in a database and is retrieved when there is a login request with the same username-password credentials.

When a request is made, credential verification is done using the database. When the password is incorrect, an unsuccessful attempt is registered in the database against that username. When the password is correct, then using the user's parameters and dataset, the probability of fraudulence is calculated, and the request is then moved onto the risk analyzer.

The probability of fraudulence (P) scaled to a hundred is calculated using machine learning algorithms like Support Vector Machines (SVM) and logistic regression. SVM is one of the best supervised machine learning algorithms. SVM works well even in higher dimensions, i.e. when the dataset has multiple attributes or parameters. In the approach presented in this paper, we need to classify the user's login requests into two classes: fraudulent and genuine, along with the probability with which the algorithm is classifying it as fraudulent. SVM is fed with the dataset containing the user's login parameters mentioned above. SVM uses a hyperplane to classify data into two different classes. A kernel function is used to convert or transform input data into the required form. A linear kernel usually classifies data with good efficiencies; if not, kernel trick can be used in SVM. A non-linear kernel will be used to transform it to higher dimensions, improving the algorithm's efficiency. SVM is programmed to indicate the probability of fraudulence which is then used by the risk analyzer to evaluate the risk of the login request.

Logistic regression is another popular machine learning algorithm that is used to classify data. Logistic regression can be used to classify the user's request as fraudulent or genuine. Logistic regression uses a decision boundary to classify data into different classes. It aims to minimize the cost function or to maximize the likelihood function. A gradient descent algorithm is used to minimize the cost function.

Along with the usage of machine learning algorithms, certain rules can be deployed to increase the authentication system's security further. For example, a rule can be deployed: If the difference between the login time is less than 10 minutes and a change in country or time zone is encountered, then an OTP verification can be made compulsory. Another rule-based example is: If more than ten unsuccessful attempts are made against the same username, physical/virtual verification by a human agent can be made compulsory, and the login process can be blocked until then. Another example is: If login time is four hours away from the usual login time, a security question can be used to challenge the user. Usually, the machine learning algorithm takes care of most instances, but as there is a possibility of misclassification, rule-based addition on top of the machine learning algorithm creates a robust authentication system.

B. Risk Analyzer

Based on the probability of fraudulence(P) scaled to hundred received by the fraudulent classifier, the risk analyzer is designed to build the risk profile for the current login request. It is classified into five classes:

Retrieval Number: 100.1/ijsce.F35080510521
DOI: 10.35940/ijsce.F3508.0710621
Journal Website: www.ijsce.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication

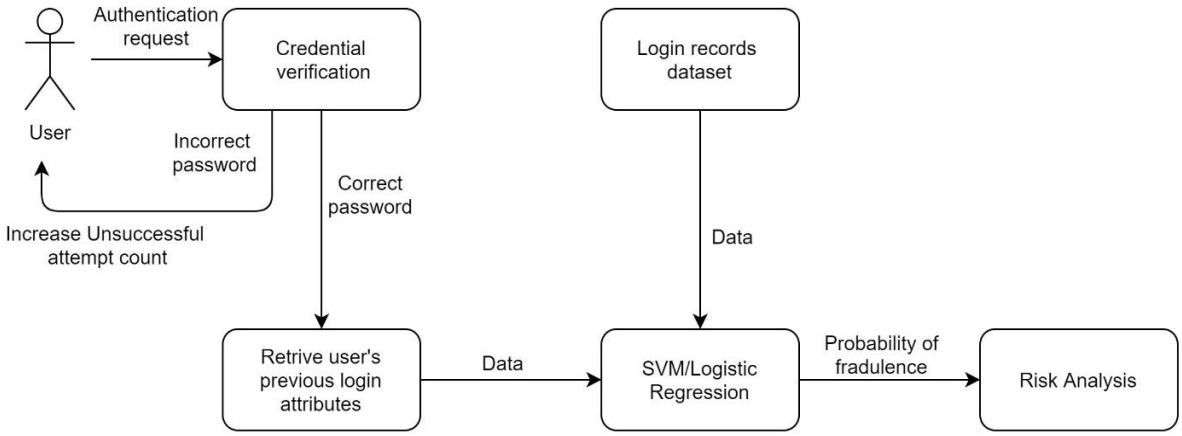

Fig. 3. Fraudulent classifier

- Level 1 Risk: 40<P<=50
- Level 2 Risk: 50<P<=60
- Level 3 Risk: 60<P<=70
- Level 4 Risk: 70<P<=80
- Level 5 Risk: P>80

Level 1 risk indicates a low-risk profile. Thus, a security question that the user previously chose can be used to verify the user's credibility. The user will be presented with an advanced risk of the next subsequent level if he cannot pass the challenge. Level 2 indicates a low to medium risk profile. The user is challenged to choose a sequence of images in a particular order that the user had chosen during the registration process, or a challenge can be presented to choose a set of images that fit into their pre-chosen category. Level 3 risk indicates a medium risk profile and thus needs to be verified with a one-time password (OTP) sent to their registered email address or phone number. Level 4 indicates medium to high risk profile. A digital signature is given as a challenge to the user. Level 5 risk indicates a high risk profile. Thus, the login request is denied for the current request. A physical/virtual verification needs to be conducted by a human agent of the organization authenticating the user to unlock the account.

IV. RESULT AND ANALYSIS

The proposed systematic approach was implemented using the Python programming language. The user needs to register with a username, password, security question, image sequence, or a category of images the user wishes to choose in the future if a challenge is presented when a risk is encountered. The machine learns along with a supervised dataset for the first few login requests. Till five login attempts are made by the user, the user will be challenged by an OTP. After five attempts, risk analysis will be done by the machine for all the subsequent requests. Table 2 indicates different user login records and the probability of fraudulence associated with it. The algorithm performed very well in all the cases. The algorithm demanded a challenge at least of risk level 2 when the user was fraudulent. The algorithm did not demand any challenge for 92% of the cases on average when the user was genuine. The last row in Table 2 indicates the case when the algorithm demanded a challenge for a genuine user, who was trying to access from a different resolution monitor, location, language, and IP address using the same device with a cookie. Such a structured level of verification will provide ease of use for the user without compromising security.

Table- I: Experimental Results

S L No	Username	Percentage of cases when the challenge was not demanded for genuine users
1	gau12	89
2	rstp17	97
3	vind1456	100
4	pri432	86
5	basu35	95
6	raks8	93
7	allen67	84

Table 1 indicates the percentage of cases when the challenge was not demanded for genuine users after five login attempts. The experiment was conducted on seven users, with fifty logins for each user. The result was tabulated only after the first five login attempts, as OTP is compulsory for the first five login attempts to keep the authentication system secure. After five login attempts, the risk analyzer can analyze the risk of the user based on their profile. It can be seen that after 20 attempts, the percentage of cases when the challenge was not demanded is fairly above 86 percent. The results also depend on the user's login parameters and their behavior. For example, the third user in Table 1 didn't have any significant changes in subsequent logins, so the third user was never challenged.

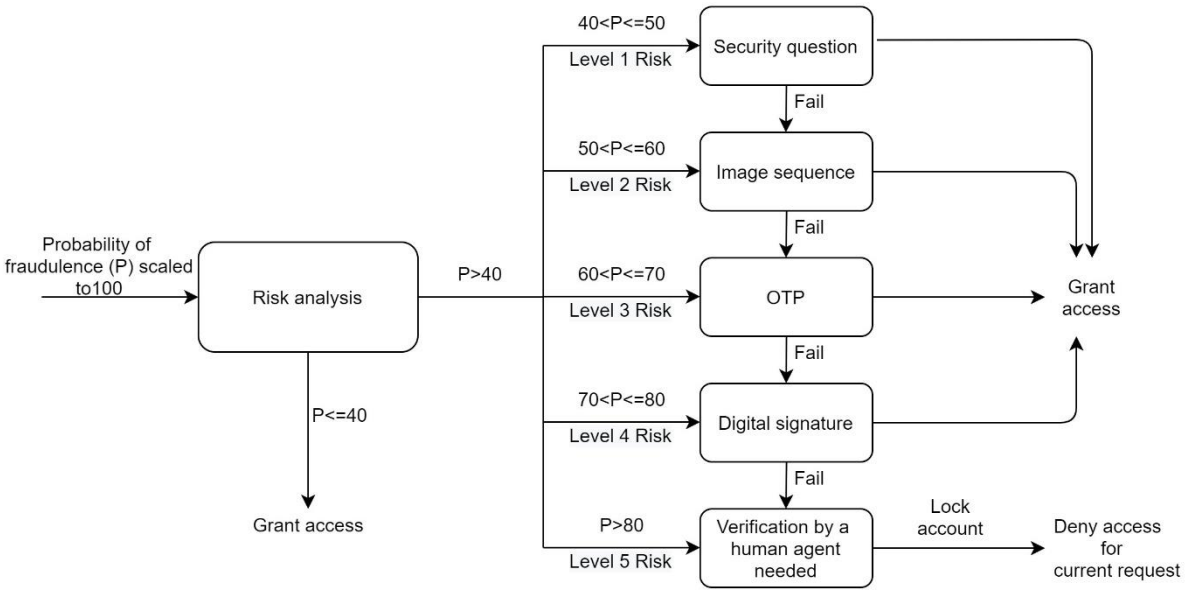

Fig. 4. Risk Analyzer

Table- II: User parameters for a particular user

Username	IP address	Availability of Cookie	Browser	OS	Unsuccessful attempts	Location	Time zone	Language	Display Resolution	Login time	P
gau12	202.1.36.22	Yes	Edge 90.0	Win 10	0	Bengaluru	IST	EN	1980*1080	21:43	22
gau12	202.1.36.22	Yes	Safari 14.0	Mac 11	0	Bengaluru	IST	EN	1980*1080	20:02	38
gau12	128.2.6.102	No	Edge 90.0	Win 10	0	New Delhi	IST	EN	1980*1080	15:17	68
gau12	195.1.33.18	No	Chrome 90.0	Win 10	2	San Jose	PST	EN	1980*1080	07:43	85
gau12	202.1.36.22	Yes	Edge 90.0	Win 10	0	Bengaluru	IST	EN	1980*1080	22:10	15
gau12	145.15.16.8	Yes	Edge 90.0	Win 10	0	Mumbai	IST	HI	3840*2160	02:56	59

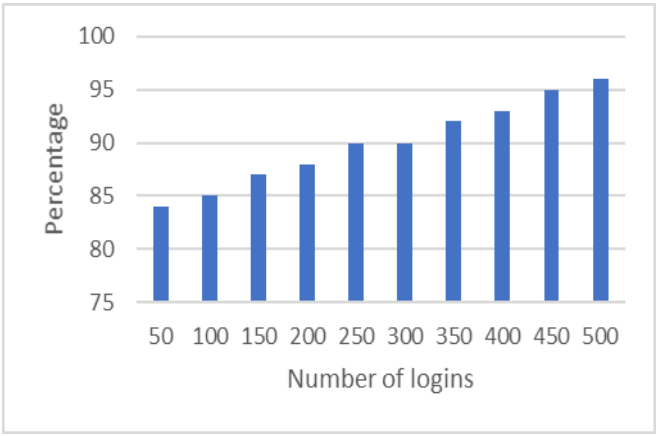

Fig. 5. The plot of the number of logins vs. percentage

Figure 5 indicates the plot between the number of logins and the percentage of cases when the challenge was not demanded for genuine users for the seventh user in Table 1. It was observed that the percentage of cases when the challenge was not demanded for genuine users kept on increasing as more and more logins were conducted by a genuine user of that username. When the user had conducted almost five hundred logins, the accuracy of the algorithm has increased to more than 95 percent.

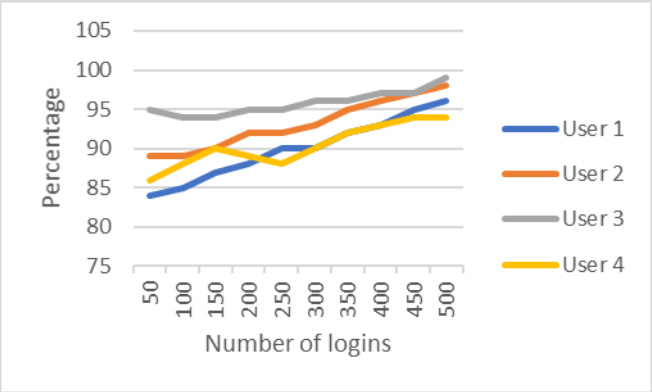

Fig. 6. The plot for different users with increasing logins

Figure 6 indicates the plot between the number of logins and the percentage of cases when the challenge was not demanded for genuine users, for different users. It is evident that over the increasing number of logins, the percentage of cases when the challenge was not demanded for genuine users keeps increasing. User 1 in the graph indicates the seventh user in Table 1, User 2 in the graph indicates the first user in Table 1, User 3 in the graph indicates the fifth user in Table 1, User 4 in the graph indicates the fourth user in Table 1. Sometimes a dip can also be noticed with the increasing number of logins, which is due to significant changes in the user parameters. But the overall trend with an increasing number of logins by the users is an increase in the percentage of cases when the challenge was not demanded for genuine users. In the whole experiment, no login attempts by fraudulent users were let through before an additional authentication.

V. CONCLUSION AND FUTURE WORK

This paper presents a systematic architecture to produce a secure authentication system using risk-based authentication and machine learning algorithms. A genuine user is not required to prove his credibility by providing multiple credentials until any risk is associated with the request; only a fraudulent user needs to pass multiple factors of authentication. This makes the login process very relaxed for a genuine user with additional security. Any variations in a user's login attributes are considered a potential risk. Based on the probability of the user being fraudulent or not genuine, a challenge is provided to the user to prove his credibility. This whole approach of secure authentication depends on the user's login parameters; classifying an imposter as a fraudulent user who uses the same device with all the same login parameters is out of the scope of this paper. Work can be taken to improve the efficiency of the classifying machine learning algorithm to deliver better and accurate results.

ACKNOWLEDGMENT

We would like to thank Rashtreeya Vidyalaya College of Engineering for providing facilities to conduct research, friends, and family who have supported us throughout. We would like to, Dr. K N Subramanya, Principal, R V College of Engineering, and Dr. Ramakanth Kumar P, Professor and Head of the Department of Computer Science, R V College of Engineering, for providing us all the required environment to conduct research.

REFERENCES

1. Wiefling, Stephan, Luigi Lo Iacono and Markus Dürmuth. “Is This Really You? An Empirical Study on Risk-Based Authentication Applied in the Wild.” ArXiv abs/2003.07622 (2019).
2. Freeman, D & Jain, Sakshi & Duermuth, Markus & Biggio, Battista & Giacinto, Giorgio. “Who Are You? A Statistical Approach to Measuring User Authenticity”. 10.14722/ndss.2016.23240 (2016).
3. D. V. Bailey, M. Durmuth, and C. Paar, “Statistics on password re-use and adaptive strength for financial accounts,” in Security and Cryptography for Networks, ser. Lecture Notes in Computer Science, vol. 8642. Springer, pp. 218–235 (2014).
4. M. Bishop and D. V. Klein, “Improving system security via proactive password checking,” Computers & Security, vol. 14, no. 3, pp. 233–249 (1995).
5. Diep N. N., S. Lee, Y. K. Lee, H.J. Lee, "Contextual Risk-based Access Control", Security and Management, pp. 406-412 (2007).
6. IP2Location - Identify Geographical Location and Proxy by IP Address. https://www.ip2location.com/.
7. Cheng P. C., P. Rohatgi, C. Keser, P. A. Karger, G. M. Wagner, and A. S. Reninger, "Fuzzy Multi-Level Security: An Experiment on Quantified Risk Adaptive Access Control", IBM Research Report RC24190 (2007).
8. Steinegger, R.H., Deckers, D., Giessler, P., Abeck, S. “Risk-based authenticator for web applications”. In: Proc. EuroPlop ’16. pp. 16:1–16:11, ACM (2016).
9. Akhtar, N., ul Haq, F. “Real time online banking fraud detection using location information”. In: Proc. CIIT 2011, pp. 770–772, Springer (2011).
10. R. O. Duda, P. E. Hart, and D. G. Stork. “Pattern Classification”. Wiley, Interscience Publication, (2000).
11. Oded Peer. Yedidya Dotan, Yael Villa and Marcelo Blatt. "USING BASELINE PROFILES IN ADAPTIVE AUTHENTICATION". US Patent 8,621,586 BI, December 31(2013).
12. Peter Chapin, Christian Skalka, and X. Sean Wang, “Risk assessment in distributed authorization”, Proceedings of the 2005 ACM workshop on Formal methods in security engineering, Fairfax, VA, USA, November 11-11 (2005).
13. Nathan Dimmock, Jean Bacon, David Ingram, and Ken Moody, “Risk models for trust-based access control (TBAC)”, In Proceedings of the Third Annual Conference on Trust Management (iTrust 2005), volume 3477 of LNCS. Springer-Verlag, May (2005).

AUTHORS PROFILE

H. A. Gautham, is a Bachelor of Engineering student at the Department of Computer Science and Engineering, R.V. College of Engineering, Bengaluru, Karnataka, India. His areas of interest are Cloud computing, Machine learning, and Big Data Analytics.

Dr. Ramakanth Kumar P, is Professor and Head of the Department at Department of Computer Science and Engineering, R.V. College of Engineering, Bengaluru, Karnataka, India. His interest lies in Digital Image Processing, Pattern Recognition, Natural Language processing.