WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Enhancing Password Security through Deep Learning and Risk-Based Authentication

Kottakota Kavyasree[1], Mr. G. Dharmaraju[2], Kotla Pramodhini[3], Koyyala Sahithi[4], Kuramana Shalini[5]

[1,3,4,5]UG Scholar, Department of CSE, GMRIT, Rajam, Andhra Pradesh, India
[2]Assistant Professor, Department of CSE, GMRIT, Rajam, Andhra Pradesh, India
kavyasree.kottakota@gmail.com[1], dharmaa.surya@gmail.com[2], kotlapramodhini@gmail.com[3], koyyalasahithi@gmail.com[4], shalinikuramana@gmail.com[5]

ABSTRACT:

Passwords remain the most widely used way to authenticate users on online sites, but because they are easy to guess and exploit, they frequently provide a serious security concern. Strong deep learning algorithms, for example, can be trained to identify and take advantage of weak passwords or ones that have been compromised in data breaches. This insightful information can be used to teach people how to create stronger passwords. Risk-based authentication (RBA), which uses sophisticated machine learning techniques to assess the possibility that a login attempt is fake, is another efficient way to improve password security. RBA enables the fast detection of any suspicious behavior and the taking of appropriate action, such as asking further authentication, to prevent illegal access. The power of RBA and deep learning combined can greatly improve password security.

Keywords: Passwords, Deep learning, Risk-Based Authentication (RBA), Machine learning

I. Introduction

In today's digital age, the vulnerability of password security has emerged as a critical concern amid the relentless evolution of cyber threats. Conventional password systems are weak points that invite unwanted access and jeopardize confidential information because they are vulnerable to dictionary and brute force attacks. A paradigm shift in the field is brought about by the combination of deep learning and risk-based authentication (RBA), as a response to this vulnerability. Using the power of deep learning techniques, particularly recurrent neural networks (RNNs) and convolutional neural networks (CNNs), this novel approach aims to transform password security. These clever models are excellent at determining the strength of passwords because they analyze user behavior, break down complexity metrics, and extrapolate conclusions from past data. The foundation for proactive protection against possible breaches is laid by their capacity to identify patterns and anomalies. To deep learning's advantage, risk-based authentication shows itself to be a powerful ally. RBA adds layers of security by dynamically modifying authentication protocols based on contextual cues such as device specs, location, and user behavior. These factors are continuously assessed by machine learning algorithms, which then assign risk scores based on which additional authentication layers are activated in the event of aberrations. This adaptive strategy foresees and reduces possible risks before they jeopardize system integrity. The combination of these innovative technologies marks the beginning of a new phase in password security. The objective of this integration is to provide a strong, user-centric authentication process, in addition to strengthening systems against frequent cyberattacks. It's a calculated joint venture that uses AI-powered insights to protect private information while keeping up with the ever-changing world of cyberattacks. This investigation delves into the unexplored domain of cybersecurity in an effort to safeguard a smooth user experience online and build a strong barrier against unwanted access.

II. Background

A. The Rise of Online Platforms and the Role of Passwords

The growth of the internet and the profusion of digital platforms have fundamentally changed the ways in which we interact, communicate, and transact. The significance of trustworthy authentication techniques cannot be overstated in this day of rapid technological development. Passwords are now the standard method for user identity verification on a variety of online platforms because they are widely used and simple to use.

B. The Vulnerabilities of Passwords

Passwords are intrinsically vulnerable to a variety of security threats, despite being widely used. Users run the risk of becoming the target of malicious actors using strategies like password guessing, brute-force attacks, and phishing because weak or easily guessable passwords are combined with the

widespread practice of using them across multiple accounts. A lot of passwords that allow unauthorized access to user accounts have also been leaked as a result of the all too frequent data breaches.

C. The Need for Enhanced Password Security

Password security needs to be improved urgently because cyber threats are constantly changing and personal data is becoming more and more valuable. It is no longer possible to protect sensitive data and protect users from ever-changing security threats using antiquated password-based authentication systems.

D. The Potential of Deep Learning and RBA

Deep learning and risk-based authentication (RBA) have come together to create a new buzz in the password security space. With the knowledge and resources provided by this state-of-the-art technology, users can tackle weak passwords head-on by creating stronger and more secure login credentials. Furthermore, RBA's machine learning implementation enables continuous evaluation of login attempts, quickly identifying any possible fraudulent activity.

E. Combining Deep Learning and RBA for a Robust Solution

Security for passwords takes on new dimensions when RBA and deep learning are combined. Deep learning algorithms enable users to create passwords that are more robust and secure. In addition, RBA serves as a watchful hound, continuously tracking login attempts and identifying possible dangers. Together, they significantly reduce the susceptibility of password-based authentication systems to cyberattacks and strengthen defenses in general.

III. Risk-based Authentication (RBA)

Risk-based Authentication (RBA) is a dynamic and adaptable method for protecting sensitive data and digital identities in the ever-evolving field of cybersecurity. RBA distinguishes itself from conventional authentication techniques by utilizing an approach to access control that is more contextual and intuitive. RBA continuously modifies its authentication procedure by assessing the risk associated with each login attempt in real time. This is also known as adaptive authentication. This thorough analysis considers a number of variables, including transaction details, location, device specs, and user behavior. RBA can effectively assess the validity of login attempts and stop possible fraud by using this information.

The fundamental strength of RBA is its capacity for ongoing risk analysis and evaluation. There are several important steps in this risk assessment process:

Data Gathering and Analysis: RBA collects and examines a vast array of data from multiple sources, such as device information, location data, transaction details, past user behavior patterns, and login attempts made by the user.

Risk Scoring: RBA rates each login attempt according to its level of risk using sophisticated machine learning algorithms. Various factors gathered during data collection are taken into account when calculating this score.

Authentication Adjustment: RBA establishes the proper level of authentication for every login attempt based on the risk score. Simple credentials, like passwords, may work for low-risk attempts. One-time passwords (OTPs) or multi-factor authentication (MFA) are examples of additional authentication factors that are necessary for higher-risk scenarios.

Constant Monitoring: RBA is a continuous process that keeps an eye on user behavior and adjusts as new risks and threats arise. This ongoing observation guarantees that sensitive data is adequately protected by the authentication process.

A new way of thinking has been brought about by risk-based authentication. RBA distinguishes itself as a potent ally in protecting sensitive data while simultaneously offering a seamless and user-friendly experience because of its adaptability to constantly shifting risk assessments. RBA's critical role in protecting digital identities and maintaining the integrity of online platforms will only increase as the landscape of cyber threats changes.

IV. Methodology

This paper presents a comprehensive approach to deep learning integration with RBA, addressing key elements including data preparation and collection, model selection and training, risk assessment, and authentication. The main goal is to smoothly incorporate deep learning models into the authentication system so that contextual data-driven adaptive authentication and timely risk assessment are possible. In order to guarantee the successful application of RBA, the methodology also gives top priority to crucial elements like user experience, explainability and transparency, and privacy and security.

Phase 1: Data Collection and Preparation

1. Identify Data Sources: Compile pertinent data sources that offer insights into login trends and user behavior. This could consist of:

- User login attempts (profitable and unprofitable)
- User device information (type, operating system, and browser information)
- User location information (IP address, geolocation)

- User behavior data (frequency, duration of logins, and device swapping)

2. Data Preprocessing: To ensure the quality and consistency of the data, clean and prepare it. This could include:

- Managing missing values
- Imputing missing data as needed
- Normalizing data to a standard scale
- Feature engineering to glean pertinent details from unprocessed data

Phase 2: Model Selection and Training

1. Choose a Deep Learning Model: Depending on the type of data and the intended result, choose a suitable deep learning model. Sequential data such as user login events are ideally modeled by long short-term memory (LSTM) networks and recurrent neural networks (RNNs).

2. Model Training: Use the prepared data to train the selected deep learning model. The process takes place in below order:

- Data splitting into training, validation, and testing sets
- Model hyperparameter optimization for optimal performance
- Model performance evaluation on the validation set and necessary adjustments

Phase 3: Risk Assessment and Authentication

1. Risk Scoring: Connect the authentication system to the trained deep learning model. The model will calculate a risk score for each login attempt based on the contextual data that has been supplied.

2. Adaptive Authentication: Apply adaptive authentication in accordance with the risk assessment. Standard authentication techniques, like usernames and passwords, might be adequate for low-risk logins. It might be necessary to take extra authentication steps (like multi-factor authentication or one-time passwords) for high-risk logins.

3. Constant Monitoring: To make sure the deep learning model is effective in identifying unusual user behavior and possible security risks, keep an eye on its performance and periodically retrain it with new data.

Additional Considerations:

- Privacy and Security: To preserve user privacy while preserving the efficacy of risk-based authentication, apply data minimization strategies and secure data handling methods.
- Explainability and Transparency: To increase user trust and provide transparency into the risk assessment process, develop explainable AI mechanisms.
- User Experience: To prevent interfering with the user's experience, strike a balance between security and convenience by only requiring extra authentication steps when absolutely necessary.

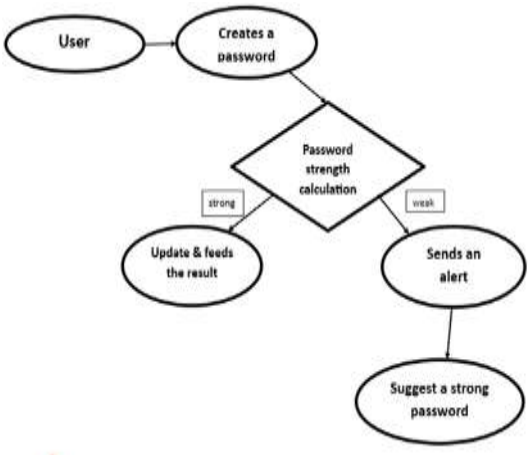

Fig-1: flowchart of the process when user creates new password

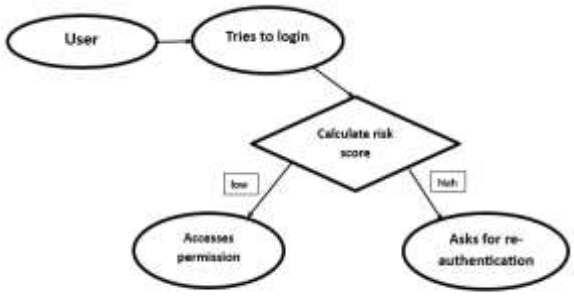

Fig-2: flowchart of the process for password risk calculation

Systems for creating passwords and detecting fraud are shown in the two images.

To find and stop fraudulent activity, the fraud detection system makes use of data collection, machine learning, risk scoring, a rules-based engine, and investigation. To find out if fraud has happened, data is collected from multiple sources, examined by human analysts, assessed by a rules-based engine, and subjected to machine learning algorithms for analysis.

In order to determine password complexity and assist users in creating strong passwords, the password creation system makes use of a password strength calculator. The password strength is calculated by the calculator using variables like length and character variety. The user receives an updated password if it is strong. If weak, a suggestion for a strong password is provided along with an alert. The purpose of this system is to protect user accounts and improve password security.

V. Results and Discussion

i. Results

S. No.	Author(s)	Method	Dataset	Accuracy	Other metric
1.	Stephan Wiefling, Paul Rene Jorgensen,Sigurd Thunem & Luigi Lo Lacono	Freeman Et AL and ML model	Login data collected at SSO service	--	--
2.	Picard, C., & Pierre, S.	RLAuth model using DRL	MPU from crawdad	--	G-mean = 92.62%
3.	Murmu, S., Kasyap, H., & Tripathy, S	Bi-GAN	MySpace & RockYou	87%	--
4.	Zhang, T., Cheng, Z., Qin, Y., Li, Q., & Shi, L.	LSTM, RNN, and GAN	Leaked password datasets	Depends on dataset used	--
5.	Hong, K. H., & Lee, B. M.	ANN	Dataset was created using Wikipedia & other sources	95.74%	--

Table-1:table for the outcomes of different methods

The outcomes of five distinct techniques for estimating the accuracy of login data are displayed in Table 1. The table's columns are:

S.No.: The method's serial number.

Author(s): The method's author(s) listed in alphabetical order.

Method: The method's name.

Dataset: The designation of the dataset that was used to assess the approach.

Accuracy: The method's accuracy as measured by the dataset.

Other metric: An additional metric, like the F1-score or G-mean, is used to assess the methodology.

The table demonstrates that, with an accuracy of 95.74% on the dataset, the ANN method is the best-performing approach. The RLAuth model, with an accuracy of 92.62%, is the second best method.

Additionally, the table demonstrates how the methods' accuracy varies based on the dataset that is used. On the Leaked password datasets, for instance, the LSTM, RNN, and GAN approach performs best; on the MySpace and Rock You datasets, on the other hand, the Bi-GAN method performs best.

In general, the table presents a valuable summary of the efficacy of five distinct techniques in forecasting the precision of login information.

Here is a more detailed explanation of each method:

(Artificial Neural Network)*: ANNs are a subset of machine learning algorithms that are applicable to a range of tasks, including classification. In this instance, given the login data, the ANN is trained to predict if a login will be successful or unsuccessful.

RLAuth model (Reinforcement Learning Authentication model)*: RLAuth model, is a kind of reinforcement learning algorithm that can be used to teach agents how to perform complicated tasks. In this instance, various login data is used to train the RLAuth model to learn how to authenticate users.

Bi-GAN (Bidirectional Generative Adversarial Network)*: Bi-GANs, are a kind of generative adversarial network that can produce data that is realistic. The Bi-GAN is trained to produce accurate login data in this instance.

GAN, RNN, and LSTM (Generative Adversarial Network, Recurrent Neural Network, and Long Short-Term Memory)*: Neural network types that can be used for a range of tasks, such as generation and classification, include LSTMs, RNNs, and GANs. Here, the accuracy of login data is predicted using the LSTM, RNN, and GAN methods on various datasets.

ii. Discussion

Since deep learning models are capable of efficiently assessing password strength, they constitute a major advancement in online security. In order to determine strength, these models take into account common patterns, character diversity, and password length. When used in password strength meters, they give users instant feedback, which motivates them to create stronger passwords and enhances overall security in digital environments. This proactive approach emphasizes the value of knowledgeable user behavior and strong security measures by representing a standard shift in cybersecurity consciousness and defense mechanisms. The use of deep learning models to assess password strength is a noteworthy development in the reinforcement of online security protocols. Strong tools for recognizing patterns and simulating sequential data in passwords are convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short-term memory (LSTM) networks. These models accurately forecast the strength of newly generated passwords because they were trained on large datasets with a range of password strengths. Deep learning models' practical application is demonstrated by their incorporation into a number of password management and security tools, including Google Password Manager, 1Password, and Have I Been Pwned? These systems use deep learning to evaluate the strength of passwords by taking into account variables like length, character variety, and the existence of recurring patterns or phrases. This program helps improve password security and teaches users how to make strong passwords, which reduces their exposure to cyberattacks. Online service providers can incorporate password strength meters and provide users with real-time feedback on how strong their passwords are by utilizing deep learning models. By encouraging users to create stronger passwords and steer clear of popular weak patterns, this proactive approach improves overall security posture across digital platforms.

VI. Conclusion

An important development in online security is the incorporation of deep learning models into risk-based authentication, password generation, and security. Different systems can assess password strength more efficiently when convolutional neural networks, recurrent neural networks, and long short-term memory networks show flexibility in identifying password patterns. By using these models to predict password strength, detect compromised passwords, and apply RBA, platforms such as Google Password Manager promote proactive user education and cyber threat prevention. These models foster informed user behavior in addition to strengthening security protocols. Deep learning models provide strong protection against unwanted access in addition to RBA's adaptive authentication process, which adapts dynamically based on contextual factors like device type and user behavior, reducing risks even with weak passwords. Deep learning and RBA together greatly improve organizational security, protecting users from cyberthreats in the constantly changing digital environment.

VII. References

1. Picard, C., & Pierre, S. (2023). RLAuth: A Risk-based Authentication System using Reinforcement Learning. IEEE Access.

2. Murmu, S., Kasyap, H., & Tripathy, S. (2022). PassMon: a technique for password generation and strength estimation. Journal of Network and Systems Management, 30, 1-23.

3. Wiefling, S., Dürmuth, M., & Lo Iacono, L. (2020, December). More than just good passwords? a study on usability and security perceptions of risk-based authentication. In Annual Computer Security Applications Conference (pp. 203-218).

4. Wiefling, S., Dürmuth, M., & Iacono, L. L. (2021). Verify it's you: How users perceive risk-based authentication. IEEE Security & Privacy, 19(6), 47-57.

5. Zhang, T., Cheng, Z., Qin, Y., Li, Q., & Shi, L. (2020, December). Deep learning for password guessing and password strength evaluation, A survey. In 2020 IEEE 19th International conference on trust, security and privacy in computing and communications (TrustCom) (pp. 1162-1166). IEEE.

6. Hong, K. H., & Lee, B. M. (2022). A deep learning-based password security evaluation model. Applied Sciences, 12(5), 2404.

7. Wu, Y., Wang, D., Zou, Y., & Huang, Z. (2022, August). Improving Deep Learning Based Password Guessing Models Using Pre-processing. In International Conference on Information and Communications Security (pp. 163-183). Cham: Springer International Publishing.

8. Tao Zhang, Zelei Cheng, Yi Qin , Qiang Li, Lin Shi (2020). Deep Learning for Password Guessing and Password Strength Evaluation. In International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom).

9. Wiefling, S., Lo Iacono, L., & Dürmuth, M. (2019). Is this really you? An empirical study on riskbased authentication applied in the wild. In ICT Systems Security and Privacy Protection: 34th IFIP TC 11 International Conference, SEC 2019, Lisbon, Portugal, June 25-27, 2019, Proceedings 34 (pp. 134-148). Springer International Publishing.

10. Acien, A., Morales, A., Vera-Rodriguez, R., Fierrez, J., & Monaco, J. V. (2020, September). TypeNet: Scaling up keystroke biometrics. In 2020 IEEE International Joint Conference on Biometrics (IJCB) (pp. 1-7). IEEE.

11. Rivera, E., Tengana, L., Solano, J., Castelblanco, A., López, C., & Ochoa, M. (2020, November). Risk-based authentication based on network latency profiling. In Proceedings of the 13th ACM Workshop on Artificial Intelligence and Security (pp. 105-115).

12. Buriro, A., Gupta, S., Yautsiukhin, A., & Crispo, B. (2021). Risk-driven behavioral biometric-based one-shot-cum-continuous user authentication scheme. Journal of Signal Processing Systems, 93, 989-1006