# Password Strength Evaluation via Zipf's Law and Password Entropy

Jiajing Zhang[1,*,#], Yang Xu[2,#], Hongda Liu[3]

[1]School of Cyber Engineering, Xidian University, Xi'an, China, 710018

[2]School of Mathematical Sciences, Soochow University, Suzhou, China, 215021

[3]College of Computer Engineering, Qingdao City University, Qingdao, China, 266106

*Corresponding author: 13902079731@163.com

[#]These authors contributed equally.

**Abstract.** Password strength assessment is pivotal for safeguarding information security. While conventional methods primarily emphasize password length and character diversity, they often overlook character distribution patterns. Our paper introduces a novel password strength evaluation method leveraging Zipf's Law and the password entropy model. Despite the importance of password strength evaluation in ensuring information security, current methods frequently rely solely on password length and character combinations, disregarding character distribution patterns. By integrating Zipf's Law, commonly observed in natural language and passwords, into our evaluation framework, we propose a more comprehensive and precise method. Through experiments utilizing diverse password datasets and comparative analysis with traditional approaches, we validate the superior accuracy and reliability of our proposed method in assessing password strength. This research holds significance in enhancing password security assessment methodologies and offers more effective support for password management and information security.

**Keywords:** Password Strength Evaluation, Zipf's Law, Password Entropy.

## 1. Introduction

In today's digitalized society, cybersecurity holds paramount importance owing to the rapid evolution of information technology. Passwords play a crucial role in account security, yet users often opt for weak passwords. This paper delves into the exploration of password strength evaluation models to furnish scientifically sound security recommendations.

In the current field of cybersecurity, password strength evaluation is both important and challenging. Wang et al. [1] developed "zxcvbn," a tool based on rules and heuristics for assessing password strength. However, its effectiveness is limited by the size of the dataset it relies on, especially since these datasets are primarily composed of historically leaked passwords, which may not keep pace with emerging, uncompromised password patterns. Moreover, due to the relatively complex computation process of "zxcvbn," it may encounter performance issues in resource-constrained environments, particularly when numerous concurrent password evaluations are required. At the same time, Herley and van Oorschot [2] proposed a model based on user behavior, but its generalizability is limited. Guo et al. [3] introduced a lightweight method that provides rapid assessment but may lack depth. Golla et al. [4] suggested a machine learning-driven approach, which requires extensive data and regular updates. The statistical method presented by de Carné de Carnavalet et al. [5] is fast but potentially insufficient. Thai et al. [6] considered a model that accounts for interdependencies between characters, requiring ample training data. Khern-am-nuai et al. [7] introduced an approach aware of user cognition that needs further validation.

The work presented in this article aims to address the limitations of the above methods by proposing a password strength evaluation model that is both rapid and highly adaptable. Our model combines statistical analysis, rule engines, and lightweight machine learning techniques to improve the accuracy and efficiency of evaluations while reducing reliance on large datasets. In this way, we

hope to provide a more flexible, efficient, and deployable solution to meet the evolving threats in cybersecurity.

Domestic scholars have also made significant contributions. Zhang et al. [8] proposed a machine learning-based password guessing algorithm. Guo et al. [9] suggested a detection method, while Yang et al. [10] proposed a population-based evaluation. Wang et al. [11], Wu et al. [12], and Lu et al. [13] explored deep learning, probability language models, and recurrent neural networks, respectively. However, there remains a need for a comprehensive evaluation method that considers multiple factors to enhance accuracy.

Our research has made contributions to password strength assessment in the following areas:

- Developed comprehensive evaluation criteria.
- Conducted an in-depth study and application of Zipf's Law, combined with the password entropy model, through extensive data analysis and model training, to establish an accurate password strength evaluation model.
- Based on the precise evaluation model established above, the methods based on Zipf's law and the entropy model are compared to verify its accuracy and applicability.

In this paper, we explore password security research based on Zipf's Law and the password entropy model. Section 2 introduces the theoretical basis of password strength assessment, including Zipf's Law, password entropy, and related methods. Section 3 provides an in-depth study of password strength assessment using Zipf's Law and the password entropy model, analyzing their application principles and effectiveness through frequency relationships. Section 4 presents the experimental design, data analysis, and results comparison between Zipf's Law and cryptographic entropy models, aiming to identify their advantages and disadvantages. Finally, Section 5 summarizes the findings, identifies areas for improvement, and discusses future directions for cryptographic security research.

## 2. Basic Concepts And Theoretical Basis

The overall structure of our paper is illustrated in Figure 1 below. First, we introduce the theoretical foundations of password strength assessment, including Zipf's Law and password entropy. Next, we analyze the principles and effectiveness of these models. This is followed by the experimental design, data analysis, and a comparison of results to identify the strengths and weaknesses of each model. Finally, we summarize our findings.

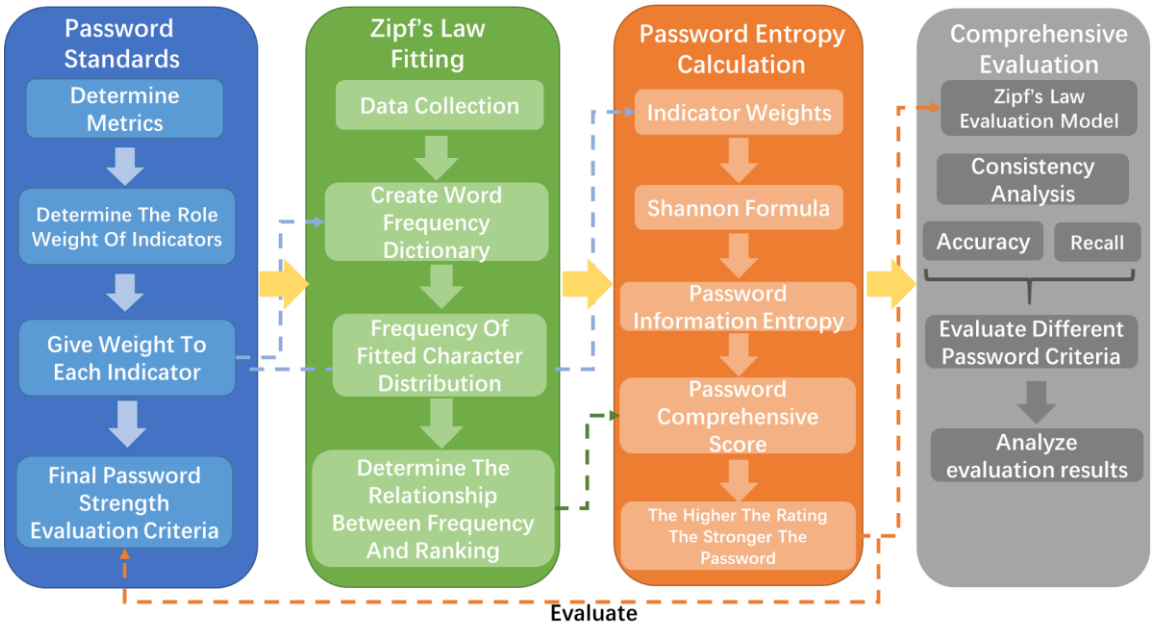


**Figure 1.** Structure Framework

### 2.1. Password Strength Assessment Overview

Identity authentication is crucial for ensuring the security of information systems, and passwords, as one of the most common authentication methods, play a pivotal role in this regard. Password strength assessment aims to evaluate the security of passwords chosen by users to prevent them from being guessed or cracked by attackers. This assessment typically considers factors such as password complexity, length, and the variety of characters included.

- **From the Attacker's perspective:**

Attackers often exploit weaknesses in passwords to carry out guessing attacks. They may analyze the system's password security mechanisms, including password generation strategies, strength assessment methods, authentication processes, and storage vulnerabilities. Additionally, attackers may leverage patterns in user password creation, such as common password structures, reuse of passwords, and the use of personal information, to enhance their guessing strategies.

- **From the Defender's perspective:**

Defenders need to design and implement effective password strength evaluation mechanisms to enhance password security. Current methods for evaluating password strength primarily include rule-based approaches, pattern detection, and attack simulation algorithms. Rule-based evaluation methods assess password strength based on factors like length and character types. Pattern detection methods evaluate password complexity by identifying common patterns in password construction. Meanwhile, evaluation methods based on attack simulation gauge password strength by simulating attacks and assessing the difficulty of successfully cracking the passwords.

### 2.2. Introduction to Zipf's Law

Zipf's Law, initially proposed by Zipf in 1949, describes the rank-frequency distribution of vocabulary in natural language. It stipulates an inverse relationship between the frequency of a word and its rank in a corpus. Specifically, in a corpus arranged in descending order of frequency, the rank and frequency of a word follow an inverse relationship, expressed as follows:

$$f_r = \frac{c}{r} \tag{1},$$

Where $C$ is a constant specific to the corpus. This implies that the frequency of the most frequent word is approximately double times that of the second most frequent word, and so on. Notably, this law has explained various phenomena such as internet topology (Faloutsos et al., 1999), US firm sizes (Axtell, 2001), and the distribution of Linux software packages (Maillart et al., 2008). By excluding less common passwords (i.e., those appearing fewer than three or five times) and employing linear regression, we observed a similar adherence to Zipf's Law in real password datasets. For a given dataset, the relationship between password rank and frequency conforms to the equation

$$f_r = \frac{C}{r^s} \tag{2},$$

Where $C$ and $s$ represent dataset-specific constants influenced by various factors, including network service types, password policies, and user demographics. Further confirmation of this consistency was provided through visualization using log-log plots, where the linear relationship between

$$\log(f_r) = \log C - s \cdot \log\ r \tag{3}.$$

### 2.3. Crypto Entropy And Its Applications In Cryptography

Entropy is a significant concept in information theory, used to quantify the uncertainty of information. In cryptography, password entropy serves as a crucial concept for evaluating the strength and unpredictability of passwords. The computation of password entropy is based on the concept of information entropy, representing a measure of uncertainty in passwords. Typically, information entropy is employed to denote password uncertainty. The password's information entropy can be computed using the Shannon entropy formula:

$$H(X) = -\sum p(x) \cdot \log_2(p(x)) \tag{4}$$

where $p(x)$ denotes the probability of a character in the password, and $x$ represents the length of the password. Entropy reflects the uncertainty of password values; higher entropy indicates greater uncertainty, making the password more difficult to crack.

### 2.4. Overview Of Related Technologies And Methods

The password strength assessment method based on Zipf's Law and entropy calculation includes the following steps: first, conducting large-scale password data collection, then performing data preprocessing and character frequency calculation, fitting the character frequency distribution to Zipf's Law, computing password entropy, utilizing a weighted multi-criteria evaluation standard, and finally calculating a comprehensive score for password strength assessment.

## 3. Establishment Of Password Strength Evaluation Standards

### 3.1. Indicator Mining

Existing password evaluation standards offer valuable metrics for assessing security and guiding users to create stronger passwords:

- **Password Entropy**: Measures complexity based on character set size and length. Higher entropy indicates better security.
- **Character Classes**: Evaluates the diversity of character types (uppercase, lowercase, numbers, special characters) to enhance complexity.
- **Length**: Longer passwords are harder to crack; standards often recommend minimum lengths to boost security.
- **Avoiding Common Passwords**: Checking for common or dictionary-based passwords reduces vulnerability to guessing attacks.
- **Password History Check:** Ensures new passwords differ from previous ones, minimizing reuse and enhancing security.

### 3.2. Password Strength Evaluation Criteria

Password Strength Score: Considers password entropy, character classes, length, avoidance of common passwords, and historical password check.

**a. Password Entropy Calculation:** Measures password complexity based on character set size and length. Formula as follow:

$$Password\ Entropy = \log_2(Character\ Set\ Size^{Password\ Length}) \tag{5}$$

**b. Character Classes Consideration:** Each additional character class adds 5 points.

**c. Password Length Evaluation:** For each additional character beyond the threshold, add 10 points.

**d. Avoiding Common Passwords:** Deduct 20 points for including common password terms. Common Passwords List: Used to check if the password is easily guessable or crackable.

**e. Historical Password Check:** Deduct 30 points if the new password's similarity with historical passwords exceeds the threshold. Exceeding this threshold indicates high similarity with historical passwords.

This evaluation standard comprehensively considers password complexity, length, character classes, avoidance of common passwords, and historical password checks. It provides users with more accurate security recommendations, thereby enhancing overall system security.

# 4. Model Introduction

## 4.1. The Application Principle Of Zipf's Law In Password Strength Assessment

Zipf's Law states that the frequency of a word in a corpus is inversely proportional to its rank. In password strength assessment, Zipf's Law can be used to evaluate password entropy, indicating its randomness. Higher entropy implies greater difficulty in cracking. Thus, we propose a method based on Zipf's Law for password strength assessment. Password entropy, which is inversely proportional to its rank, denotes stronger passwords with higher entropy values. By analyzing word frequency and ranking, we calculate a password's entropy to assess its strength.

## 4.2. Algorithm Design Of Zipf's Law In Password Strength Assessment

The algorithm design of Zipf's Law in password strength assessment includes the following steps:

- **Create Word Frequency Dictionary:**Gather password data, including common password dictionaries and real user passwords. Common password dictionaries consist of publicly available lists featuring frequently used passwords and common English word combinations. Real user passwords involve a subset of authentic passwords anonymized for privacy.
- **Calculate Password Entropy:**Utilize Zipf's Law to compute the frequency and ranking of each word in the password and derive its entropy. Break down the password into words or characters, tally the frequency of each, and rank them accordingly. Employ the entropy formula from information theory to gauge the password's randomness and complexity.
- **Experimental Analysis:**Assess the password entropy value experimentally to evaluate its strength and compare it with traditional methods. By computing the entropy value, classify passwords into strong, medium, and weak categories based on distinct entropy value intervals.

## 4.3. Concept And Basic Principles Of Cryptographic Entropy Model

Password entropy, a measure of password complexity and randomness, is crucial for assessing password strength. This section introduces the password entropy model and its key principles.

Password entropy gauges the uncertainty or randomness within a password, determined by the diversity of characters and password length. Higher entropy implies greater randomness and complexity, bolstering resistance against brute force attacks. Typically measured in bits, it indicates the minimum bits needed to encode password information.

Key principles:

1. **Entropy formula** $H$ **:** $p_i$ denotes the probability of the $i$ th character in the password, and $N$ represents the character set size. Entropy formula as follow:

$$H = -\sum_{i=1}^{n} p_i \log_2(p_i) \tag{6}$$

2. **Character set diversity impacts entropy:** A broader set, comprising uppercase, lowercase, numbers, and symbols, elevates entropy. Here, $N$ represents the character set size, $n$ denotes the number of unique characters, $p_i$ represents the probability of each character, and $N$ signifies the character set size.

3. **Password length:** Longer passwords yield higher entropy due to the increased number of possible combinations, thereby enhancing security. The impact of password length on entropy can be quantified by the formula:

$$\Delta H = \log_2(N) \tag{7},$$

where $\Delta H$ is the entropy increase after adding a character, and $N$ is the character set size.

4. **Uniform distribution:** Ideally, characters should be evenly distributed in a password, ensuring each has an equal probability. Non-uniform distribution may reduce entropy, compromising security.

# 5. Experimental Design And Result Analysis

## 5.1. Our Experimental Design

We conducted experimental verification using datasets from Yahoo, Baidu, and others to validate the superior performance of Password Strength Evaluation via Zipf's Law and Password Entropy(PSE-ZLPE). For detailed experimental code, please refer to https://github.com/HondaRUM/password-strength-evaluation.

Initially, we utilized Zipf's Law distribution to validate password strength across various datasets. Employing diverse datasets, including common password datasets and natural language datasets like the Wishbone.io Dataset, CSDN, Facebook, and Yahoo, we computed occurrence frequencies and analyzed the occurrence frequencies of passwords. Data closely following Zipf's Law distribution indicated stronger passwords, while significant deviations suggested weaknesses. According to the experimental results, the evaluation score can be accurate to 2 decimal places, which is more fine-grained and accurate than other evaluation models that can only describe the degree of strengthThe experimental results demonstrated the accuracy of PSE-ZLPE in assessing password strength across different datasets. The experimental results of the PSE-ZLPE model are shown in Figure 2 below.

Next, based on the password dataset, we quantified password complexity and randomness by calculating password entropy. We then comprehensively evaluated password strength by combining the distribution characteristics of Zipf's Law and the password entropy calculation results. Passwords fitting Zipf's Law distribution well and having higher password entropy were considered strong passwords, while those with less fitting and lower entropy were deemed weak passwords. We determined the relevant weights, with the final result being the optimal weight: Zipf distribution weight = 0.35, password entropy weight = 0.1, validated on various password datasets. The experimental results depicted the sensitivity of PSE-ZLPE, showcasing its comprehensive and accurate evaluation of password strength.

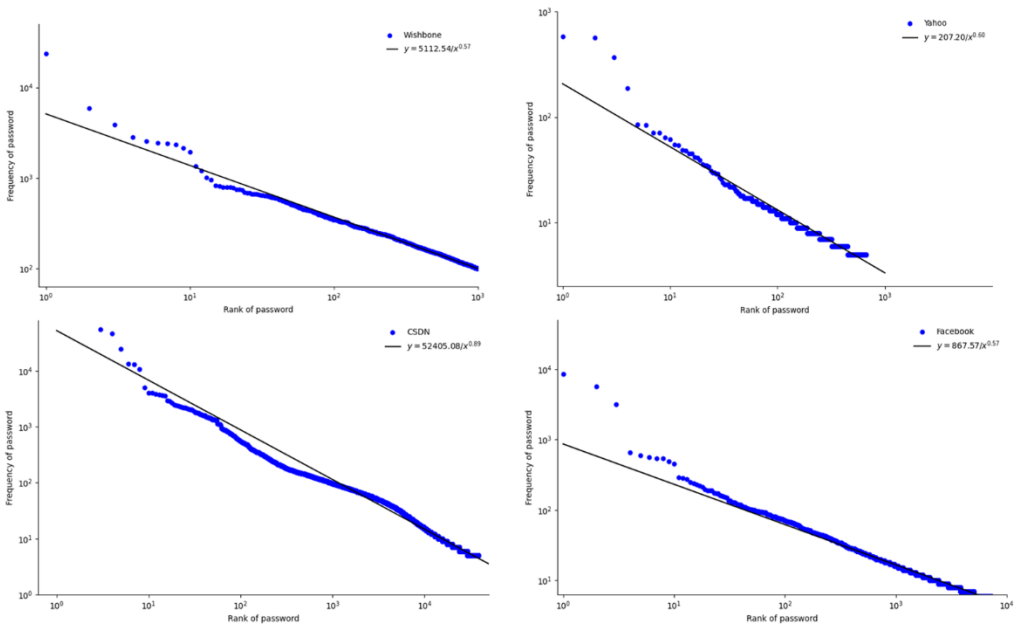


**Figure 2.** Experimental results based on PSE-ZLPE model[1]

We utilized the derived weights to construct a password scoring system capable of assessing various datasets with diverse passwords. Among these datasets, CSDN represents a Chinese website, while the others are international platforms. The password dataset comprises numeric sequences, Chinese and international catchphrases, alphanumeric combinations (including digits, uppercase and lowercase letters, and a minimum of two special characters), and randomly generated strings. We categorized scores from 0 to 50 as weak, 50 to 80 as moderate, and 80 to 100 as strong. Notably,

randomly generated passwords like "12!@asAS69~" and "W9puO097hw6gf" are typically found on educational websites due to their complexity, making them difficult to remember. The scores of different website passwords are shown in Table 1 below.

**Table 1.** Password ratings for different websites

| | **CSDN** | **Yahoo** | **Wishbone** | **Facebook** |
|---|---|---|---|---|
| 12345678 | 47.14 | 45.03 | 36.34 | 40.95 |
| iloveyou | 43.47 | 42.13 | 46.73 | 38.32 |
| woaini520 | 44.28 | 43.65 | 45.74 | 46.07 |
| Zhang123 | 47.14 | 45.02 | 46.83 | 47.95 |
| babygirl200 | 63.02 | 65.42 | 68.32 | 59.50 |
| phone83035 | 71.42 | 59.47 | 79.52 | 54.09 |
| Franny123469 | 88.84 | 87.88 | 86.49 | 80.84 |
| Calendersucks19 | 93.47 | **100** | 93.23 | 94.72 |
| 12!@asAS69~ | 90.60 | 92.78 | 95.46 | 95.93 |
| W9puO097hw6gf | **100** | 97.42 | **96.85** | **98.87** |

**We rate 0-50 as weak, 50-80 as strong, and 80-100 as strong.**

Through the aforementioned experiments and validations, PSE-ZLPE assumes a pivotal role in password strength assessment:

1. Comprehensive Evaluation: By integrating Zipf's Law distribution and password entropy, PSE-ZLPE comprehensively evaluates password strength, considering both the commonality and complexity of passwords simultaneously.

2. Weighted Password Attributes: Assigning distinct weights to different password attributes enables a nuanced evaluation, with weights adjusted based on the actual characteristics of the password. At the weight calculated by our programming, the effect of zipf’s law fits best.As the zipf is a natural law with wide recognition, the better the zipf fit, the more closely results in password strength evaluation outcomes aligned with real-world scenarios.

### 5.2. Comparative Experiment with Existing Website Password Strength Assessment Standards

In the current network security field, password strength assessment is both important and challenging. GitHub now is using zxcvbn password strength estimator which is developed by Dropbox and is a rule-based and heuristic tool for assessing the strength of passwords.

Key Method:

**1.Data preprocessing:** It collects and processes large amounts of password data, including leaked passwords and common dictionaries. These data are used to construct the basis for dictionary matching and pattern recognition.

**2.Pattern recognition:** Common patterns in passwords are identified using regular expressions and heuristic algorithms. Common patterns include:

- Continuous characters: such as "123456" or "abcdef".
- Duplicate characters: aaaaaa or 111111.
- Keyboard mode: such as qwerty or asdfgh.

**3.Dictionary match:**Match the entered password to the pre-built dictionary. The password is considered less intense if it is in the dictionary.zxcvbn Use multiple dictionaries, including common phrases, names, popular culture references, etc.

**4.Entropy calculation:** Use the Shannon entropy formula in information theory

**5.Comprehensive score:**The results of pattern recognition, dictionary matching, and entropy calculation are considered comprehensively. The score is divided into five grades: very weak, weak, medium, strong and very strong.

By comparing the two password evaluation methods with the actual output effect, we can get some disadvantages of zxcvbn password strength estimator compared with the proposed PSE-ZLPE model.

The sources and quantities of the cipher suites we selected are shown in Table 2 below. We choose the strength rating of the GitHub website as a reference, as shown in Figure 3 below.

**1.Dependent data sets:** The effectiveness of zxcvbn depends on the size and quality of its dictionary and pattern library. For new password patterns not included in the dictionary, the assessment results may be inaccurate. With the advent of new cryptographic patterns and attack methods, zxcvbn password strength estimator needs to regularly update its dictionary and rule base to maintain its validity and accuracy.

**2.Complexity limitations:** Since zxcvbn password strength estimator uses complex pattern recognition and matching algorithms, it may encounter performance problems in resource-constrained environments, especially when a large number of concurrent password evaluation is required. The PSE-ZLPE model used in this paper has low requirements for system environment resources, just running the Python program to obtain results.

**3. Fuzzy interval:** zxcvbn password strength estimator only gives a rough evaluation range of whether it is safe or not, which is more specific than PSE-ZLPE model, and the evaluation effect is slightly vague.

**Table 2.** Source and number of selected password sets[3]

| Source | CSDN | Yahoo | Wishbone | Facebook |
|---|---|---|---|---|
| Number | 4510000 | 301552 | 7656356 | 1734218 |

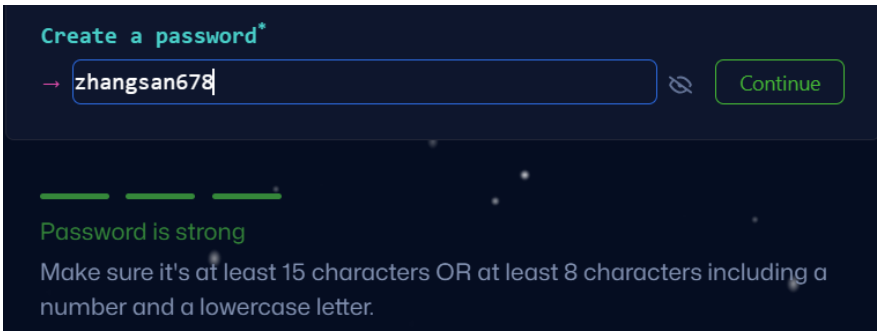


**Figure 3.** Strength ratings of GitHub websites[7]

## 6. Conclusion

In conclusion, our experimental results demonstrate that the PSE-ZLPE model offers a superior, comprehensive evaluation of password strength by integrating Zipf's Law distribution and entropy calculations. This model provides finer granularity and greater accuracy compared to existing methods like zxcvbn. Additionally, PSE-ZLPE is more adaptable to new password patterns and requires fewer system resources, making it a valuable tool for reliable and precise password strength assessment in the evolving landscape of cybersecurity.

## References

[1] Wang D, Shan X, Dong Q, et al. No single silver bullet: Measuring the accuracy of password strength meters[C].32nd USENIX Security Symposium (USENIX Security 23). 2023: 947-964.

[2] Amador J, Ma Y, Hasama S, et al. Prospects for improving password selection[C]//Nineteenth Symposium on Usable Privacy and Security (SOUPS 2023). 2023: 263-282.

[3] Guo Y, Zhang Z. LPSE: Lightweight password-strength estimation for password meters[J]. computers & security, 2018, 73: 507-518.

[4] Golla M, Dürmuth M. On the accuracy of password strength meters[C]//Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. 2018: 1567-1582.

[5] Asaduzzaman A, D'souza D, Uddin M R, et al. Increase Security by Analyzing Password Strength using Machine Learning[C]//2024 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON). IEEE, 2024: 32-37.

[6] Thai B L T, Tanaka H. A statistical Markov-based password strength meter[J]. Internet of Things, 2024, 25: 101057.

[7] Khern-am-nuai W, Hashim M J, Pinsonneault A, et al. Augmenting password strength meter design using the elaboration likelihood model: Evidence from randomized experiments[J]. Information Systems Research, 2023, 34(1): 157-177.XI.

[8] Zhang Zijian. Research on Password Guessing Algorithm Based on Classical Machine Learning[D]. Peking University, 2019. DOI: 10.26929/d.cnki.gbeju.2019.000041.

[9] Guo Anchao. Research and Implementation of Password Strength Detection Based on Active Information Collection[D]. Beijing University of Posts and Telecommunications, 2021. DOI: 10.26969/d.cnki.gbydu.2021.002490.

[10] Yang Xiao. Research on Password Strength Evaluation Based on Population[D]. East China Normal University, 2019.

[11] Wu Yu. Research on Password Strength Evaluation and Password Enhancement Based on Probabilistic Language Model[D]. East China Normal University, 2018.

[12] Bonneau, Joseph. "The science of guessing: analyzing an anonymized corpus of 70 million passwords." IEEE Security & Privacy 12.6 (2014): 52-57.

[13] Florêncio, Dinei, and Cormac Herley. "A large-scale study of web password habits." Proceedings of the 16th international conference on World Wide Web. 2007.

[14] Dourado, Andrei, and Artur Ziviani. "On the Entropy of Real-World Passwords." IEEE Transactions on Information Forensics and Security 13.9 (2018): 2181-2193.