Evaluating the Strength and Availability of Multilingual Passphrase Authentication

Chi-en Amy Tai
University of Waterloo
amy.tai@uwaterloo.ca

Urs Hengartner
University of Waterloo
urs.hengartner@uwaterloo.ca

Alexander Wong
University of Waterloo
alexander.wong@uwaterloo.ca

***Abstract*—Passwords are a ubiquitous form of authentication that is still present for many online services and platforms. Researchers have measured password creation policies for a multitude of websites and studied password creation behaviour for users who speak various languages. Evidence shows that limiting all users to alphanumeric characters and select special characters resulted in weaker passwords for certain demographics. However, password creation policies still concentrate on only alphanumeric characters and focus on increasing the length of passwords rather than the diversity of potential characters in the password. With the recent recommendation towards passphrases, further concerns arise pertaining to the potential consequences of not being inclusive in password creation. Previous work studying multilingual passphrase policies that combined English and African languages showed that multilingual passphrases are more user-friendly and also more difficult to guess than a passphrase based on a single language. However, their work only studied passphrases based on standard alphanumeric characters. In this paper, we measure the password strength of using a multilingual passphrase that contains characters outside of the standard alphanumeric characters and assess the availability of such multilingual passwords for websites with free account creation in the Tranco top 50 list and the Semrush top 20 websites in China list. We find that password strength meters like zxcvbn and MultiPSM surprisingly struggle with correctly assessing the strength of non-English-only passphrases with MultiPSM encountering an encoding issue with non-alphanumeric characters. In addition, we find that half of all tested valid websites accept multilingual passphrases but three websites struggled in general due to imposing a maximum password character limitation.**

I. Introduction

Passwords are still a critical component for users to authenticate on websites [1]. To improve the formation of stronger passwords, many websites often employ password creation policies based on the US National Institute of Standards and Technology (NIST) guidelines [2]–[4]. In the latest released guideline, NIST placed greater emphasis on password length rather than complexity (i.e., adding special characters) [5]. Subsequently, passphrases have become the recommended best practice for users as they are easier to remember and are often longer than a randomly derived password [6].

Interestingly, research studying the creation behaviour for different languages revealed that cultural differences can lead to differences in the structure and strength of created passwords [7], [8]. In some instances, language barriers can also result in weaker passwords from certain demographics [9]. For example, Bonneau [9] reported that users who spoke German and Korean chose the strongest passwords whereas users who spoke Indonesian chose the weakest passwords. Subsequently, the new popular recommendation to use passphrases raises concerns as this would also detriment certain non-English-speaking groups.

Motivated by these findings, Maoneke et al. [10] studied the juxtaposition of substrings from multiple languages to improve passphrase security. Specifically, they evaluated the strength of multilingual passphrase policies that combined both English and African languages [10]. They illustrated that multilingual passphrases are more user-friendly and more difficult to guess than a passphrase based on a single language [10]. However, their work only studied passphrases based on standard alphanumeric characters [10].

Building on their findings, we expand their work and measure the password strength of using a multilingual passphrase that contains characters outside of the standard alphanumeric characters. Given that the top languages in the world are English (1.5 B), Mandarin Chinese (1.1 B), and Hindi (608.8 M) [11], we present an analysis of multilingual passphrase authentication availability that focuses on combining English and Chinese. Using entropy, we create three categories of strong passphrases: (1) English-only, (2) Chinese-only, and (3) Multilingual with both English and Chinese. We assess and compare the strength of these passphrases using zxcvbn and MultiPSM against the computed entropy values. Furthermore, we also assess the availability of such multilingual passwords for websites with free account creation in the Tranco top 50 list [12] and the Semrush top 20 websites in China list [13]. We discover that both zxcvbn and MultiPSM incorrectly report the strength of multilingual passphrases and find similar results to Bonneau and Xu [14] with 12 of the evaluated websites fully supporting multilingual passphrases.

II. Related Work

A. Password Differences Across Languages

A survey of 409 people in Hungary and Serbia that looked at their password habits for smart devices showed differences in password behaviour between the two countries [15]. Notably, a greater proportion of people in Serbia use personal data and

Workshop on Measurements, Attacks, and Defenses for the Web (MADWeb) 2025
28 February 2025, San Diego, CA, USA
ISBN 979-8-9894372-9-0
https://dx.doi.org/10.14722/madweb.2025.23004
www.ndss-symposium.org

meaningful words in their passwords whereas more people in Hungary prefer using more complex passwords (different case letters, numbers, special characters) [15]. When studying 79,760 passwords leaked from a Middle Eastern bank, AlSabah et al. [16] also noticed cultural differences present in passwords between four demographic groups (Arabic speakers, India and Pakistan, Philippines, and English speakers) that would inflate their presumed strength. For example, compared to the other three groups, it was almost twice as likely for India and Pakistan passwords to include a special character [16]. Additionally, English passwords more frequently included "!" whereas passwords from the Philippines contained "_" more often [16]. The user's name and birth year were also seen to be more frequent in passwords from India, Pakistan, and the Philippines, but their analysis showed that English users actually had a higher guessability than the other groups based on adversarial models [16].

Surprisingly, when Bonneau [9] analyzed roughly 70 million passwords from Yahoo! users, they found that English- and Chinese-speaking users had passwords that were similar in strength [9]. However, when Li et al. [17] reviewed a dataset of 100 million leaked passwords from various Chinese and international websites and took a deeper look at the specific passwords created by Chinese- and English-speaking users, they found differences in the composed passwords from both groups. Specifically, they documented that Chinese-speaking users prefer digits whereas English-speaking users prefer lowercase letters and Chinese-speaking users tend to use Chinese Pinyins whereas English-speaking users use English words [17]. Building on their research, Han et al. [18] concluded that while passwords from Chinese-speaking users may appear strong according to password meters designed for English-speaking users, these passwords could actually be significantly weaker if password guessing algorithms take into account the regional differences in password structure. This finding is also supported by Wang et al. [7] in their analysis of 73.1 million real-world Chinese web passwords. Likewise, when Chae et al. [19] analyzed over 39 million breached authentication data points and focused on breached Korean email and password data, they discovered that Korean passwords often had Korean names and frequently occurring Korean word representations, which may also reduce the strength of Korean passwords beyond what was previously reported. As shown in these research works, there are password differences across languages that may impair the actual strength of the chosen passwords given cultural similarities and patterns.

B. Measuring Passphrase Strength in Offline Attacks

To measure passphrase strength in offline attacks, it is important to consider the evolution of password strength measurement given the recency of passphrases compared to passwords and the similarity of passphrases and passwords.

Entropy was first introduced by Shannon [20], who applied it to analyze English text [21]. Entropy has since been adapted for use in information theory to measure the randomness of passwords and passphrases. Massey [22] termed guessing entropy as the lower bound for the estimated average number of successive guesses for passwords. Building on this concept, Yan et al. leveraged entropy for password checking with the suggestion to filter passwords with low entropy [23]. Then, Komanduri et al. [24] used entropy to measure password predictability and compared the calculated entropy for different password composition policies to provide recommendations to improve password strength whilst maintaining user usability.

On the other hand, Ma et al. [25] specifically investigated password entropy and password quality and concluded that password entropy inadequately measures password quality as it assumes an all-or-nothing nature. To expand, the authors dislike how all passwords of the same length have the same amount of entropy and find it an unsuitable measurement [25]. Instead, the authors propose a new scheme termed password quality indicator that considers the password cracking strategy of trying obvious passwords such as those based on dictionary words [26]. They argue that password quality should consider the password's difference to dictionary words and the size of possible password characters in addition to the password length [26]. Based on their proposed scheme, Ma et al. recommended that good quality passwords have "at least 8 characters long, with at least 3 special characters plus other alphanumeric characters" [26]. However, current research has shown that these types of password composition policies are actually less effective than they seem and pose significant usability issues for users [27] rendering Ma et al.'s proposed password quality indicator inappropriate for passphrase strength measurement.

Further research on the topic of password strength measurement led to the development of multiple password strength meters that are imprecise and incoherent with each another [28]. Subsequently, MultiPSM [29] focused on combining multiple methods including Markov chains and a blocklist score for deriving the final strength. Wang et al. [30] measured the accuracy of password strength meters and showed that in offline scenarios, MultiPSM obtained the best results for Chinese and English guessing scenarios with zxcvbn also performing well compared to the other tested password strength meters. Similarly, Maoneke et al. [10] also leveraged guess numbers to assess passphrase strength and used zxcvbn [31] for pattern matching and conservative estimation. Recently, Mukherjee et al. [32] proposed a framework to systematically generate memorable and secure passphrases in English and measured the strength of their passphrases with guessrank using the min auto approach computed from the Carnegie Mellon Password Guessability service introduced by Ur et al [33]. However, the Guessability service was only designed with the English language in mind [33] rendering it ineffective for multilingual passphrase strength estimation. Subsequently, we leverage entropy to derive strong passphrases and consider MultiPSM and zxcvbn as passphrase strength measurement tools given their superior performance in prior research studies. We chose to omit the Guessability service given its inherent reliability on the English language.

C. Potential of Multilingual Passphrases

Rao et al. [34] calculated the search space for English passphrases; however, incorporating multilingual passphrases could expand this search space, potentially enhancing their security. Motivated by this thinking, Maoneke et al. [10] conducted a study with 224 university students in Southern Africa to generate passwords and experimented with enforcing a multilingual passphrase policy [10]. Using Probabilistic Context-Free Grammar (PCFG), they showed that short English language-oriented passwords were easier to guess than short passwords based on the African language and much weaker than multilingual passphrases generated from enforcing a multilingual passphrase policy [10]. However, the Latin alphabet is used as the basis for the majority of African languages and subsequently, their analysis is limited to multilingual passphrases based on alphanumeric characters [10].

Their focus on alphanumeric characters is understandable given the character encoding issues that Bonneau and Xu reported for web passwords [14]. In their 2012 analysis of 24 websites, Bonneau and Xu discovered that numerous large websites, such as Google, Amazon, and Youku, do not support non-ASCII passwords [14]. This lack of support could be due to artifacts from the history of character encoding development, which has led to a single Chinese character being expanded into larger different bytes based on the type of encoding (e.g., GB2312, UTF-8, ISO 8859-1) and potential conflicts in encoding based on the browser used for transmission, making it difficult to impose length limits and reliably validate passwords [14]. In their case studies examining leaked data sets of English, Chinese, Hebrew, and Spanish speaking users, they found that most users relied on transliterating non-ASCII passwords to ASCII (e.g., through Pinyin), changed their keyboard mappings, and used passwords based on a geometric keyboard pattern or chose only numbers [14]. However, at the time of their study, seven of the tested sites were able to correctly support non-ASCII characters for authentication and the authors had disclosed problems to the other websites as well [14]. Given the universal shift towards using UTF-8 for character encoding, it is possible that some of the previously unsupported sites are now supportive of non-ASCII passwords [14].

III. Methodology

For this study, we focus on three categories of passphrases: English-only, Chinese-only, and Multilingual (combination of English and Chinese terms).

A. Creation of Strong Passphrases

We created passphrases using a dictionary of common words from FrequencyWords [35]. For both English and Chinese, we used the full list from 2018 with 1,656,996 tokens in the English list (en) and 766,612 tokens in the Chinese list (zh_cn). Combining both the English and Chinese lists and removing duplicates results in a total of 2,423,608 tokens.

We adapt the entropy definition for passphrases and argue that it is a more representative measure of passphrase strength. Respectively, we treat each dictionary word as a single token rather than each character and consider the length of the dictionary in our computation of entropy. We assume the threat model of the adversary is that they know which dictionary we use to derive the passphrases (the data and distribution for creation). Subsequently, for passphrases chosen from two dictionaries (i.e., due to different languages), we combine the tokens from both dictionaries and remove duplicates to compute the total dictionary length.

To ensure that strong passphrases were created for experimentation, we computed the number of terms needed to generate 80 bytes of entropy for security. We chose 80 bytes based on suggestions from online security websites [36]–[39]. Using the entropy equation $E = log_2(R^L)$ [21] where E symbolizes the entropy, R denotes the range of available tokens and L symbolizes the password length, we compute the corresponding password length for a given entropy. We chose an entropy of 80 bytes, and leveraging the corresponding number of tokens in the respective FrequencyWords list, we obtain approximately 3.87 tokens for English only, 4.09 tokens for Chinese only, and 3.77 tokens for multilingual. For standardization, we use 4 tokens to create the passphrases for all three categories resulting in roughly 82.64 bytes of entropy for English only, 78.19 bytes of entropy for Chinese only, and 84.83 bytes of entropy for the multilingual category. Randomly choosing four terms from the dictionary, we create the following strong passphrases for the three categories (each term separated by "/"):

1) English-only: influential/author/typically/rethink
2) Chinese-only: 影響/作者/通常/重新
3) Multilingual: 影響/author/通常/rethink

B. Measuring Passphrase Strength

In this study, we also assessed the strength of the three created passphrases based on offline password attacks and compared the reports from two techniques: (1) zxcvbn [31] that uses pattern matching and conservative estimation adapted for the custom dictionaries in the previous section, and (2) MultiPSM [29] that combines multiple methods including Markov chains and a blocklist score. We chose these two strength measurement tools as they were shown to obtain the best results in the offline scenarios by Wang et al. [30]. Specifically MultiPSM obtained the best results for Chinese and English guessing scenarios with zxcvbn also performing well compared to the other tested password strength meters [30]. We also chose zxcvbn because it can be modified for the Chinese language and accounts for the dictionary creation nature of passphrases. Using these algorithms, we compared the strengths of the created English-only passphrase, Chinese-only passphrase, and multilingual passphrase with both English and Chinese from the previous section against the calculated entropy. Furthermore, in the process to assess the availability of multilingual authentication, some websites also present strength meters for the inputted passwords (akamai.net, apple.com, feishu.cn, github.com, and google.com). These

values are noted and also included in the comparison of passphrase strength.

C. Availability of Multilingual Authentication

To assess the availability of multilingual authentication, we evaluated all websites with free account creation on the Tranco top 50 list [12] and the Semrush top 20 websites in China list [13]. Combining the state of both lists on October 19, 2024 and removing duplicates yielded a total of 65 websites (listed in Figure 1).

google.com	gtld-servers.net	wordpress.org
amazonaws.com	googletagmanager.com	sharepoint.com
microsoft.com	googlevideo.com	t-msedge.net
facebook.com	akadns.net	youtu.be
akamai.net	windowsupdate.com	github.com
root-servers.net	microsoftonline.com	domaincontrol.com
apple.com	doubleclick.net	aaplimg.com
a-msedge.net	amazon.com	netflix.com
youtube.com	fbcdn.net	whatsapp.net
azure.com	googleusercontent.com	pinterest.com
googleapis.com	trafficmanager.net	baidu.com
akamaiedge.net	wikipedia.org	bilibili.com
twitter.com	bing.com	zhihu.com
cloudflare.com	mail.ru	qq.com (tencent)
instagram.com	l-msedge.net	csdn.net
gstatic.com	apple-dns.net	weibo.com
office.com	office.net	taobao.com
linkedin.com	fastly.net	google.com.hk
live.com	googlesyndication.com	douyin.com
tiktokcdn.com	icloud.com	163.com
1688.com	feishu.cn	douban.com
jd.com	tmall.com	

Fig. 1: Consolidated list of the websites from combining the Tranco top 50 list [12] and the top 20 websites in China published by Semrush [13] obtained on October 19, 2024.

Searching for the password policies of different websites is ineffective for our analysis as the documentation may be outdated and most websites only document password policies for length and complexity requirements rather than whether they accept non-ASCII characters (e.g., Facebook's public password policy [40]). Hence, for this analysis, we assessed the website's password policy through creating an account and attempting to change its password. As such, we assumed that the password policy employed at account creation is the same as at the change password state. We attempted to automate the procedure but given the higher restrictions on account creation and CAPTCHA requirements, we opted to manually evaluate these sites. Notably, most websites require an email to create an account and/or email verification and subsequently, the gmail created from the first tested website (google.com) was kept throughout the entire testing process and only deleted at the very end of the study. In addition, for websites that required phone number verification, a Canadian phone number was used as validation. Websites that required payment for account creation were noted, but no financial transaction was conducted and an account was not created on these websites. Instead, an attempt to check their handling of the passphrase categories through the password creation flow was made (if possible) by inputting the passphrase in the password text box and seeing if an error was generated immediately or when the next step button was clicked. In cases where account creation is not intuitive, a Google search was used to decipher how to create an account on the website (if it was possible). Some websites were initially unreachable and led to 404 error pages. For unreachable websites, a Google search was conducted to try and find the latest website link. Once the website was successfully located, we followed the general procedure detailed below with "[initial]" indicating the first passphrase that was attempted for that step. The other passphrase variations are only attempted in cases where the initial passphrase is not accepted. The task of trying to authenticate with a trimmed version of the passphrase is to ensure that the website can also properly authenticate the passphrase (on top of allowing it to be changed).

1) Create an account with only English passphrase, trying each of the following in their respective order until successful account creation. Note any minimum limit, number, and special character requirements.
   - influentialauthortypicallyrethink [initial]
   - influentialAuthortypicallyrethink
   - influential2Authortypicallyrethink
   - influential2Authortypicallyrethink@
2) Log into the account to see if successful.
3) Change the password to only Chinese passphrase, trying each of the following in their respective order until successful password change. Note any issues that occur. If the password cannot be changed, go to Step 5.
   - 影響作者通常重新 [initial]
   - 影響2作者通常重新
   - 影響2作者通常重新@
4) Log out and log into the account using the updated password to see if it is successful. If so, log out and log back in with a trimmed version of the updated password (remove the last character). Note if the login was successful with the trimmed version.
5) Change the password to use both English and Chinese passphrase trying each of the following in their respective order until successful password change. Note any issues that occur. If the password cannot be changed, go to Step 7.
   - 影響author通常rethink [initial]
   - 影響Author通常rethink
   - 影響2Author通常rethink
   - 影響2Author通常rethink@
6) Log out and log into the account using the updated password to see if it is successful. If so, log out and log back in with a trimmed version of the updated password (remove the last character). Note if the login was successful with the trimmed version.

7) Delete the account at the end (for ethical considerations to avoid having fake accounts on multiple websites).

IV. Results

A. Passphrase Strength Measurement

Despite the calculated entropy being the highest for multilingual passphrases and relatively similar for the other passphrase categories, results from the adapted zxcvbn with the custom dictionary suggested that the English-only passphrase has the highest strength. In addition, the Chinese-only passphrase is shown to have the lowest strength according to the zxcvbn results seen in Table I. Notably, it is possible that the lower result is due to the smaller dictionary size of Chinese ($\sim$766k) compared to English ($\sim$1.66M). On the other hand, all three passphrase categories have a cracking time display of centuries and both the English-only and multilingual passphrases have a zxcvbn score of 4, the highest possible score that indicates that they are very unguessable. As seen in Figure 2, for MultiPSM, the English-only passphrase is also shown to have a high password score of 9.85, but both the Chinese-only and multilingual passphrases caused an error with MultiPSM. Subsequently, even the best password strength meters (as tested in Wang et al. [7]) were unable to accurately measure the actual strength of non-English-only passphrases where the actual strength is defined through the entropy calculation.

Surprisingly, Table II shows that for the five websites with strength feedback, only the English-only passphrase was consistently considered strong. However, two of the sites (feishu.cn and google.com) had errors for the Chinese-only and multilingual passphrases. The other three websites showed a high strength for the multilingual passphrase but reduced strength for the Chinese-only passphrase. Based on a comparison of the different strengths reported by the websites, none of the five websites used the exact same strategy nor reporting display. However, the similarity between akamai.net, apple.com, and github.com suggests that they may be using the same or similar underlying backend strategy. A similar statement could also be made for feishu.cn and google.com, which both errored on the non-English-only passphrase categories.

B. Availability of Multilingual Authentication

As seen in Table III, over half of all valid sites allow for multilingual passphrases. Eight of the valid sites (facebook.com, twitter.com, cloudflare.com, instagram.com, linkedin.com, amazon.com, wikipedia.org, and netflix.com) allow for all three passphrase categories. They also handled the password authentication properly and did not grant access for the trimmed incorrect version of the passphrases. However, four of the valid sites did not allow for the Chinese-only passphrase. For akamai.net, the estimated password strength was considered too weak and thus, not accepted by the website. For fastly.net and apple.com, the Chinese-only passphrase caused an error as it needed a lowercase and uppercase letter. However, fastly.net accepted the multilingual passphrase that contained both English and Chinese terms. For github.com, an error occurred with the Chinese-only passphrase as it did not meet either option: (1) it needed at least 15 characters but the current maximum of the Chinese-only passphrase variation was deemed as length 10, and (2) it needed a lowercase letter but the Chinese-only passphrase did not have a lowercase letter.

Three sites (tiktokcdn.com, baidu.com, and 1688.com) had issues with all three tested passphrases. Interestingly, baidu.com also explicitly forbid Chinese characters. Surprisingly, apple.com had an estimated strength meter and none of the initial passphrases had met the 100% strength requirement due to their requirement of a number, a special character, and a mixture of letter casing. Password requirements for each of the three sites and apple.com are shown in Figure 3.

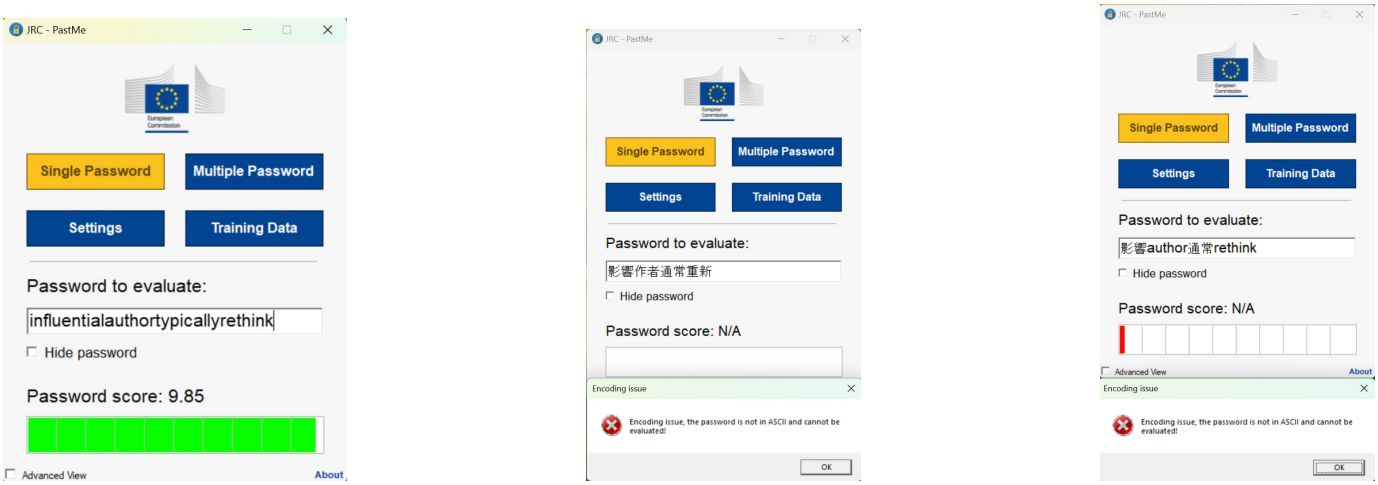
Notably five sites (google.com, amazonaws.com, microsoft.com, pinterest.com, and feishu.cn) only allow for English passphrases and forbid character types outside of uppercase letters, lowercase letters, numbers, and specific symbols. Interestingly, despite the initial similarity between amazon.com and amazonaws.com, they had distinct login flows and different results in that amazon.com allowed for Chinese characters but amazonaws.com did not.

As seen in Figure 4, the majority of sites accepted the initial passphrase but some sites only accepted variations. For the first passphrase category of English-only, one site required at least one capital letter while two sites required a number and four sites also needed the symbol in the passphrase for the password to be accepted. In the second passphrase category of Chinese-only, all sites that accepted the Chinese-only passphrase also accepted the initial passphrase except for one site that also required the number. For the third passphrase category of multilingual, all sites except four accepted the initial multilingual passphrase. One of the four was accepted after adding a number but the remaining three needed a symbol in order for the password to be accepted.

Unfortunately, most of the tested websites end up being invalid with unreachable websites, no account creation flow, account creation issues, and duplicate sites. Some of the unreachable websites include tracking sites such as apple-dns.net and t-msedge.net, and nonexistent brands like gstatic.com, and aaplimg.com. Six websites had account creation flow issues where there was no process to sign up for an account on their website. Curiously, all account creation issues came from the Semrush top 20 websites in China list with errors due to needing a phone number from Asia to register or requirements to be a Chinese citizen. Numerous domains on the list are also considered duplicate sites such as youtube.com and azure.com, which uses the same account flow as google.com and microsoft.com, respectively. Concerningly, as seen in Figure 5, the majority of duplicate sites redirect to the root service of Google and Microsoft, which only accept English passphrases. On the other hand, this presents an impactful area for improvement as updating the authentication for the root service of Google and Microsoft would also benefit multiple site experiences in the Tranco top 50 and the Semrush top 20 websites in China list.

TABLE I: Results from adapted zxcvbn with the custom dictionary where score ranges from 0 (too guessable) to 4 (very unguessable) and crack times are based on online throttling with a rate limit of 100 per hour showing that the English-only category has the highest overall passphrase strength based on zxcvbn compared to the calculated entropy.

Category	Calculated Entropy	Estimated Guesses	Guesses Log 10	Crack Time (seconds)	Score
(1) influentialauthortypicallyrethink	82.64	2.11E+17	17.3	7.60E+18	4
(2) 影響作者通常重新	78.19	1.00E+08	8.0	3.60E+09	2
(3) 影響author通常rethink	84.33	9.23E+12	13.0	3.32E+14	4

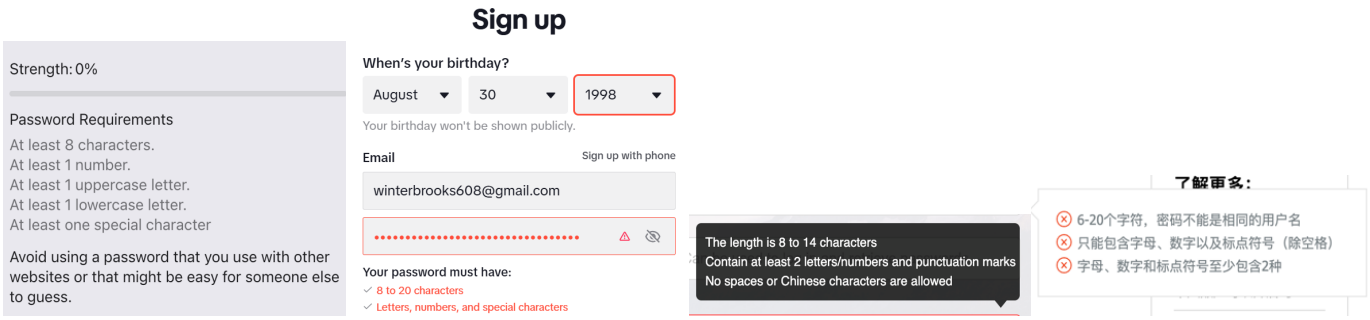


(a) English-only passphrase (b) Chinese-only passphrase (c) Multilingual passphrase

Fig. 2: Result from running the graphical application JRC-PaStMe that implements MultiPSM [41] on the English passphrase, Chinese passphrase, and multilingual passphrase with the latter two resulting in errors.

TABLE II: Best strength values reported for websites assessed for multilingual authentication.

Website	English-only Passphrase	Chinese-only Passphrase	Multilingual Passphrase
akamai.net	Good	Fair	Good
apple.com	100%	60%	100%
feishu.cn	3 bars (Strong)	Error	Error
github.com	3 bars (Strong)	2 bars (needs number and lowercase letter)	3 bars (Strong)
google.com	Strong	Error	Error



(a) apple.com (b) tiktokcdn.com (c) baidu.com (d) 1688.com

Fig. 3: Password requirements for websites with issues on all tested passphrases showcasing their inability to support passphrases given their upper limit with 20 characters and restriction on Chinese characters.

V. Discussion

A. Multilingual Passphrase Strength

Although the computed entropy for the initial passphrases were relatively similar, the two strength measurement tools of zxcvbn and MultiPSM produced contradictory results or resulted in an error. More specifically, zxcvbn showed a much weaker strength for the Chinese-only passphrase in comparison to the other two categories and favoured the English-only passphrase over the multilingual passphrase despite the entropy being the highest for the multilingual passphrase. A

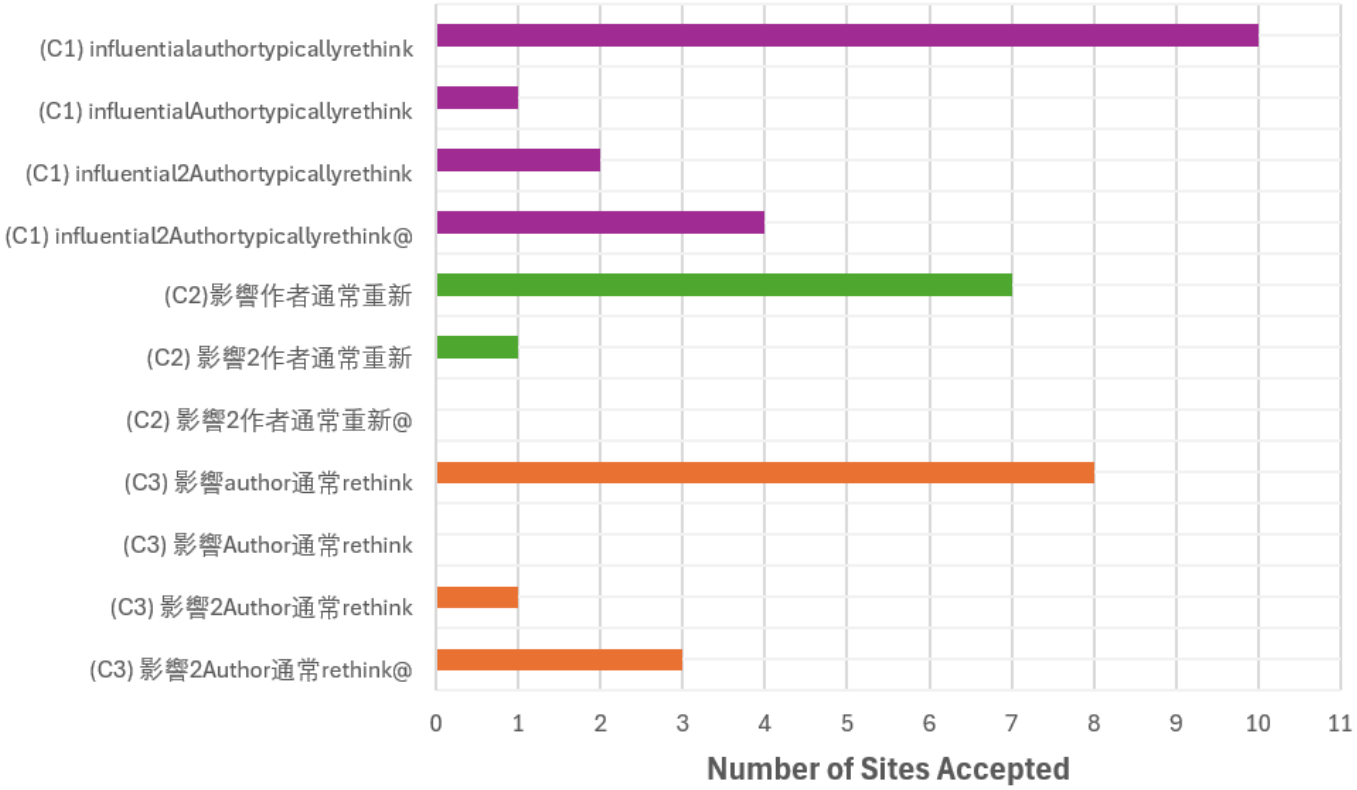


Fig. 4: Passphrases accepted by each website ordered by their order of attempt with the prefix of each passphrase indicated by the category number where C1 refers to the English-only, C2 refers to the Chinese-only, and C3 refers to the multilingual passphrase category.

TABLE III: Overview of passphrase availability for all 65 websites with over half of all valid sites allowing multilingual passphrases.

Description	Number of Sites
Valid Sites	**20**
All 3 Passphrase Categories Valid	8
English and Multilingual Passphrase Only	4
English Passphrase Only	5
Issue with Passphrase	3
Invalid Sites	**45**
Website Unreachable	12
No Account Creation Flow	6
Account Creation Issue	7
Duplicate Site	20

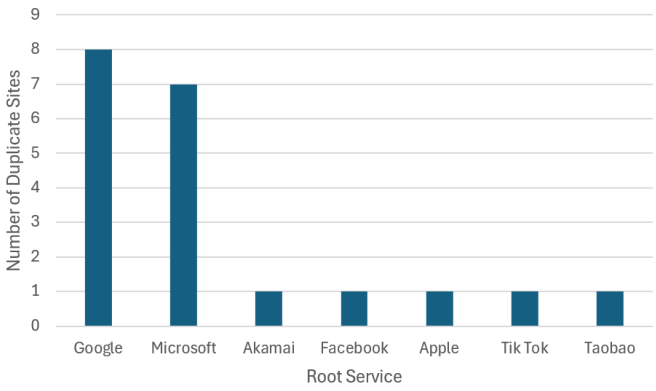


Fig. 5: Distribution of the redirected root service for the duplicate sites with the majority coming from Google and Microsoft, which only accept English passphrases.

potential explanation for the lower strength for the Chinese-only passphrase could be due to the smaller dictionary size of Chinese (about 766K) compared to English (about 1.66M). However, it is unclear why zxcvbn would report a higher strength for English over multilingual as the dictionary size for multilingual exceeds that of just English. On the other hand, MultiPSM showcased a major limitation as it produced an error when we attempted to measure the Chinese-only or multilingual passphrases. As seen in the source code, the open-source tool explicitly checks for alphanumeric characters and renders an error upon detecting any non-standard alphanumeric characters. Similarly, the strength values from the five websites that provided feedback during account creation and password change states showed a similar distribution as two out of five of the sites also had an error with the non-English-only passphrases. Additionally, the other three sites showed a much weaker strength for the Chinese-only passphrase similar to zxcvbn and only reported strong values for the English-only and multilingual passphrases. That being said, their highest

strength values were derived from a variation of the original passphrase that incorporated letters, symbols, and mixed casing. This suggests that these website password strength meters were created based on outdated documentation and that the meters are unable to accurately assess passphrase strength. Subsequently, greater effort needs to be made in production settings to improve the password strength meters to first correctly report the strength of passphrases in accordance to research findings and for researchers to derive better strength measurement systems that incorporate non-English characters in their computation of password strength.

B. Multilingual Passphrase Availability

Bonneau and Xu previously investigated character encoding issues for web passwords in 2012 for 24 websites and reported that numerous large websites, such as Google, Amazon, and Youku, do not support non-ASCII passwords [14]. In comparing their findings for websites that fully supported non-ASCII characters and websites that had policies against non-ASCII characters with our results that allowed for multilingual passphrases and forbid Chinese characters (Figure 6), there was some overlap between the sites with the majority of differences due to the sites that were explored. For example, Bonneau and Xu looked at Yahoo but this website was not in the Tranco top 50 list nor the Semrush top 20 websites in China list so we did not investigate it. The main difference was Amazon however as Bonneau and Xu found that Amazon did not support non-ASCII characters in 2012 but we found that it was possible now to create a password with Chinese characters (non-ASCII characters). However, the issue still existed on the AmazonAWS website. Curiously, the ubiquitous companies of Google and Microsoft that did not support non-ASCII characters in 2012 continue to restrict Chinese-based passwords over a decade later.

When comparing the performance between websites listed in the Semrush top 20 websites in China list against the Tranco top 50 list, it appears that the websites in the Tranco list more often have better support of multilingual passphrases. In particular, from the Semrush top 20 websites, only six sites were valid and only one of the six (Amazon.com) supported all three types of passphrases. On the other hand, from the Tranco list, there were 17 valid sites and eight of them accepted all three types of passphrases. Notably, there was an overlap between the Tranco and China list of five sites but only three were valid (Amazon.com, Github.com, Google.com) with the other two being duplicate sites.

VI. Limitations

A limitation of this work is the relatively small sample size of examined websites. Despite the consolidated list totaling 65 websites, numerous websites were considered duplicate sites as they rerouted to another site or used the exact same authentication flow as another site. In addition, a considerable number of these websites were also unreachable or did not have an account creation flow. Furthermore, for some sites on the Semrush top 20 websites in China list, a restriction was placed to require a phone number from Asia or be a Chinese citizen rendering their investigation infeasible and subsequently limiting the generalizability of this study. We do note that we were able to successfully examine 20 websites, a similar figure to previous studies in this topic [14].

Another limitation of this work is its manual nature. Considerable time had to be spent navigating each website to properly identify the sign up account creation form and change password forms, which added an overhead cost for examining each additional website. In addition, the manual procedure leaves room for potential human error and mistakes in recording results. Given the increasing presence of CAPTCHA and complexity of account creation flows, however, it was infeasible to automate the testing flow.

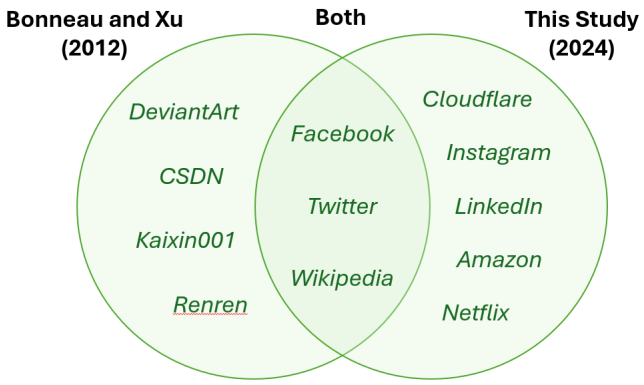
The websites studied in this work were also from the Tranco top 50 list and the Semrush top 20 websites in China list, which adds bias to the findings as it is possible that the availability of accepting multilingual passphrases could be more prevalent in other websites and across different regions. In addition, none of the studied websites were financial or banking sites, which could have drastically different handling for passwords. We also limit our investigation to websites that allow for free account creation and it is possible that paid accounts would offer multilingual passphrases more often as they could have more money invested in the storage of non-ASCII characters.

This work also only studied multilingual passphrase in the context of the Chinese language. It is possible that websites might restrict Chinese characters but accept passphrases from other languages that use non-English characters. Even so, we chose Chinese given that it is the top second language in the world and we consider the probability of websites restricting only Chinese characters but accepting other non-English characters low given the similarity of how non-English characters are handled. Thus, while the focus was on Chinese, it is reasonable to assume that the findings could be applicable to other languages with non-Latin characters, such as Hindi, Japanese, Korean, or Arabic. The underlying principles of passphrase security and multilingual support are likely to hold across these languages as well.

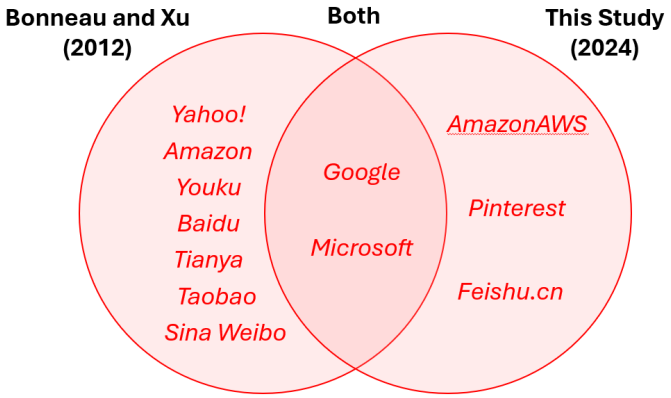
Lastly, only three major password strength assessment strategies were studied in this paper (entropy, zxcvbn, and MultiPSM). Though we did include results from websites that reported the strength during the testing procedure, it is possible that there does exist a better strength measurement tool beyond the ones that were considered in this study.

VII. Conclusion

In this paper, we evaluated the strength and availability of multilingual passphrase authentication by creating strong passphrases for three categories of passphrases: English-only, Chinese-only, and multilingual passphrases. We created strong passphrases consisting of 4 terms to obtain an entropy of roughly 80 bytes and evaluated their strength using zxcvbn and MultiPSM. We then attempted the account creation and password change process using these passphrases for a total



(a) Supporting Websites



(b) Websites that Had Policies Against

Fig. 6: Comparison of the websites from Bonneau and Xu in 2012 [14] and our study that showed some similarity with discrepancies due to differences in websites explored and a main change for Amazon that did not support non-ASCII characters in 2012 but allows for Chinese characters in multilingual passphrases now.

of 20 websites from the Tranco top 50 list and the Semrush top 20 websites in China list. Unfortunately, both zxcvbn and MultiPSM struggled with accurately reporting the strength of multilingual passphrases with the latter throwing an encoding issue upon attempting to evaluate the Chinese-only and multilingual passphrase. In terms of availability, we found similar results to Bonneau and Xu with 12 websites fully supporting multilingual passphrases and four of these sites having issues with Chinese-only passphrases due to the estimated weak strength. We also found three websites having issues with passphrases in general due to their maximum password character limit. We recommend further research to adapt password strength meters for passphrases and for multilingual character types. We also advise websites to improve their availability of multilingual passphrases and follow the lead of sites like Facebook and Amazon to improve the inclusivity of online services and platforms for all demographics.

VIII. Future Work

Future work includes expanding the list of studied websites to include those outside of the Tranco top 50 list and the Semrush top 20 websites, along with comparing different sectors of websites like financial with social media to see if there are differences in the acceptance of multilingual passphrases. Another avenue for future work would be to automate the process of account creation and password changing that leverages machine learning and scripting for reduced human intervention. Expansion of this work to other multilingual passphrases such as combining English and Hindi or combining more than two languages (e.g., English, Chinese, and Hindi) is another direction for future work. Lastly, more research could be conducted to explore how to better measure passphrase strength and incorporate the dictionaries and behaviours of multiple languages for multilingual passphrase strength meters.

Acknowledgment

This work was supported by NSERC Discovery Grant RGPIN-2020-04722.

References

[1] Gartner Research, "Craft a simple, effective password policy," https://www.gartner.com/en/documents/4687299, 2023, [Accessed 17-10-2024].

[2] K. Lee, S. Sjöberg, and A. Narayanan, "Password policies of most top websites fail to follow best practices," in *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*. Boston, MA: USENIX Association, Aug. 2022, pp. 561–580. [Online]. Available: https://www.usenix.org/conference/soups2022/presentation/lee

[3] S. A. Roomi and F. Li, "A Large-Scale measurement of website login policies," in *32nd USENIX Security Symposium (USENIX Security 23)*. Anaheim, CA: USENIX Association, Aug. 2023, pp. 2061–2078. [Online]. Available: https://www.usenix.org/conference/usenixsecurity23/presentation/al-roomi

[4] S. Alroomi and F. Li, "Measuring website password creation policies at scale," in *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 3108–3122. [Online]. Available: https://doi.org/10.1145/3576915.3623156

[5] National Institute of Standards and Technology, "Nist special publication 800-63b," https://pages.nist.gov/800-63-3/sp800-63b.html, 2023, [Accessed 17-10-2024].

[6] Government of Canada, "Best practices for passphrases and passwords (itsap.30.032)," https://www.cyber.gc.ca/en/guidance/best-practices-passphrases-and-passwords-itsap30032, 2024, [Accessed 17-10-2024].

[7] D. Wang, P. Wang, D. He, and Y. Tian, "Birthday, name and bifacial-security: Understanding passwords of chinese web users," in *28th USENIX Security Symposium (USENIX Security 19)*. Santa Clara, CA: USENIX Association, Aug. 2019, pp. 1537–1555. [Online]. Available: https://www.usenix.org/conference/usenixsecurity19/presentation/wang-ding

[8] D. Wang, Z. Zhang, P. Wang, J. Yan, and X. Huang, "Targeted online password guessing: An underestimated threat," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 1242–1254. [Online]. Available: https://doi.org/10.1145/2976749.2978339

[9] J. Bonneau, "The science of guessing: analyzing an anonymized corpus of 70 million passwords," in *2012 IEEE symposium on security and privacy*, IEEE. San Francisco, CA, USA: IEEE, 2012, pp. 538–552.

[10] P. B. Maoneke, S. Flowerday, and N. Isabirye, "Evaluating the strength of a multilingual passphrase policy," *Computers & Security*, vol. 92, p. 101746, 2020.

[11] Ethnologue, "What are the top 200 most spoken languages?" https://www.ethnologue.com/insights/ethnologue200/, 2024, [Accessed 14-10-2024].

[12] V. Le Pochat, T. Van Goethem, S. Tajalizadehkhoob, M. Korczyński, and W. Joosen, "Tranco: A research-oriented top sites ranking hardened against manipulation," in *Proceedings of the 26th Annual Network and Distributed System Security Symposium*, ser. NDSS 2019. San Diego, CA: Network and Distributed System Security (NDSS) Symposium, Feb. 2019.

[13] Semrush, "China leading websites by total visits 2024," https://www.statista.com/statistics/1456466/most-visited-websites-china/, 2024, accessed: 2024-10-19.

[14] J. Bonneau and R. Xu, "Of contraseñas, sysmawt, and mìmă: Character encoding issues for web passwords," in *Web 2.0 Security & Privacy*. San Francisco, CA, USA: IEEE, May 2012. [Online]. Available: https://www.ieee-security.org/TC/W2SP/2012/papers/w2sp12-final7.pdf

[15] D. Mandic, G. Kiss, and Z. Rajnai, "Password usage among users of smart devices in hungary and serbia," in *2024 IEEE 18th International Symposium on Applied Computational Intelligence and Informatics (SACI)*, vol. 3. Timisoara, Romania: IEEE, 2024, pp. 309–314.

[16] M. AlSabah, G. Oligeri, and R. Riley, "Your culture is in your password: An analysis of a demographically-diverse password dataset," *Computers & Security*, vol. 77, pp. 427–441, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167404818302979

[17] Z. Li, W. Han, and W. Xu, "A Large-Scale empirical analysis of chinese web passwords," in *23rd USENIX Security Symposium (USENIX Security 14)*. San Diego, CA: USENIX Association, Aug. 2014, pp. 559–574. [Online]. Available: https://www.usenix.org/conference/usenixsecurity14/technical-sessions/presentation/li_zhigong

[18] W. Han, Z. Li, L. Yuan, and W. Xu, "Regional patterns and vulnerability analysis of chinese web passwords," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 2, pp. 258–272, 2016.

[19] M. Chae, W. Shin, S. Jung, J. Yeom, D. Jeon, and H. Kim, "The threat of password guessing attacks exploiting linguistic characteristics: A case study on the korean domains," in *2024 Silicon Valley Cybersecurity Conference (SVCC)*. Los Alamitos, CA, USA: IEEE Computer Society, Jun 2024, pp. 1–4. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/SVCC61185.2024.10637306

[20] C. E. Shannon, "A mathematical theory of communication," *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[21] C. Shannon, "Prediction and entropy of printed english," *The Bell System Technical Journal*, vol. 30, no. 1, pp. 50–64, 1951.

[22] J. Massey, "Guessing and entropy," in *Proceedings of 1994 IEEE International Symposium on Information Theory*, 1994, pp. 204–.

[23] J. J. Yan, "A note on proactive password checking," in *Proceedings of the 2001 Workshop on New Security Paradigms*, ser. NSPW '01. New York, NY, USA: Association for Computing Machinery, 2001, p. 127–135. [Online]. Available: https://doi.org/10.1145/508171.508194

[24] S. Komanduri, R. Shay, P. G. Kelley, M. L. Mazurek, L. Bauer, N. Christin, L. F. Cranor, and S. Egelman, "Of passwords and people: measuring the effect of password-composition policies," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11. New York, NY, USA: Association for Computing Machinery, 2011, p. 2595–2604. [Online]. Available: https://doi.org/10.1145/1978942.1979321

[25] W. Ma, J. Campbell, D. Tran, and D. Kleeman, "Password entropy and password quality," in *2010 Fourth International Conference on Network and System Security*, 2010, pp. 583–587.

[26] W. Ma, J. Campbell, D. Tran, and D. Kleeman, "A conceptual framework for assessing password quality," 2007. [Online]. Available: https://api.semanticscholar.org/CorpusID:14233761

[27] P. G. Kelley, S. Komanduri, M. L. Mazurek, R. Shay, T. Vidas, L. Bauer, N. Christin, L. F. Cranor, and J. Lopez, "Guess again (and again and again): Measuring password strength by simulating password-cracking algorithms," in *2012 IEEE Symposium on Security and Privacy*, 2012, pp. 523–537.

[28] M. Golla and M. Dürmuth, "On the accuracy of password strength meters," in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1567–1582. [Online]. Available: https://doi-org.proxy.lib.uwaterloo.ca/10.1145/3243734.3243769

[29] J. Galbally, I. Coisel, and I. Sanchez, "A new multimodal approach for password strength estimation—part i: Theory and algorithms," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 12, pp. 2829–2844, 2017.

[30] D. Wang, X. Shan, Q. Dong, Y. Shen, and C. Jia, "No single silver bullet: Measuring the accuracy of password strength meters," in *32nd USENIX Security Symposium (USENIX Security 23)*. Anaheim, CA: USENIX Association, Aug. 2023, pp. 947–964. [Online]. Available: https://www.usenix.org/conference/usenixsecurity23/presentation/wang-ding-silver-bullet

[31] D. L. Wheeler, "zxcvbn: Low-Budget password strength estimation," in *25th USENIX Security Symposium (USENIX Security 16)*. Austin, TX: USENIX Association, Aug. 2016, pp. 157–173. [Online]. Available: https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/wheeler

[32] A. Mukherjee, K. Murali, S. K. Jha, N. Ganguly, R. Chatterjee, and M. Mondal, "Mascara: Systematically generating memorable and secure passphrases," in *Proceedings of the 2023 ACM Asia Conference on Computer and Communications Security*, ser. ASIA CCS '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 524–538. [Online]. Available: https://doi.org/10.1145/3579856.3582839

[33] B. Ur, S. M. Segreti, L. Bauer, N. Christin, L. F. Cranor, S. Komanduri, D. Kurilova, M. L. Mazurek, W. Melicher, and R. Shay, "Measuring Real-World accuracies and biases in modeling password guessability," in *24th USENIX Security Symposium (USENIX Security 15)*. Washington, D.C.: USENIX Association, Aug. 2015, pp. 463–481. [Online]. Available: https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/ur

[34] A. Rao, B. Jha, and G. Kini, "Effect of grammar on security of long passwords," in *Proceedings of the Third ACM Conference on Data and Application Security and Privacy*, ser. CODASPY '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 317–324. [Online]. Available: https://doi.org/10.1145/2435349.2435395

[35] hermitdave, "Frequencywords," https://github.com/hermitdave/FrequencyWords/tree/master, 2024, [Accessed 19-11-2024].

[36] A. Trevino, "Password entropy: What it is and why it's important," https://www.keepersecurity.com/blog/2024/03/04/password-entropy-what-it-is-and-why-its-important/, 2024, [Accessed 19-11-2024].

[37] A. Szczepanek, "Password entropy calculator," https://www.omnicalculator.com/other/password-entropy, 2024, [Accessed 19-11-2024].

[38] okta, "Password entropy: The value of unpredictable passwords," https://www.okta.com/identity-101/password-entropy/, 2024, [Accessed 19-11-2024].

[39] D. Pleacher, "Calculating password entropy," https://www.pleacher.com/mp/mlessons/algebra/entropy.html, 2024, [Accessed 19-11-2024].

[40] Facebook, "Login & password," https://www.facebook.com/help/434017203434612, 2024, [Accessed 19-10-2024].

[41] jrcpastme, "Jrc-pastme - license," https://github.com/ec-jrc/jrcpastme, 2019, [Accessed 19-11-2024].