Exploratory Analysis of Password and Login Security Methods

Sofia Huang
William & Mary

Exploratory Analysis of Password and Login Security Methods

Sofia Huang, College of William & Mary

Research mentor:
Saltuk B. Karahan, Ph.D. Lecturer and Interim Director, School of Cybersecurity

Table of Contents

1. Introduction

In 2022, the average cost of a data breach is $4.35 million. Passwords are the first line of defense against unauthorized access to accounts and important data. Ensuring that accounts are secure is vital to protect personal and corporate information. Although big tech companies are researching ways to implement password-free authentication systems, they will likely not be replaced by new technology all of a sudden. Therefore, methods to strengthen passwords and optimize security must be used until more advanced options are easily accessible. In this paper, common forms of password attacks and various methods of password security evaluation and login authentication will be discussed and analyzed.

2. Password Attacks

There are various methods that an attacker can use to gain unauthorized access to an account. The most common one is a brute force attack. This is when the attacker generates an exhaustive list of all character combinations within a password length range to try and guess the password through trial and error. Although it can be a time consuming operation, it is a simple and well-known method of gaining unauthorized access to accounts and is widely used. The reason why this method works is due to the use of weak passwords that are easy to guess and poor password etiquette, such as using the same password for multiple accounts. Many people also use passwords that are associated with their personal lives such as a pet name or a birthday, this is known as a targeted-guessing attack, which can be easily guessed if the attacker does some investigating.

There are some options for avoiding brute force attacks. One is to apply rate limiting. This limits the amount of API calls a user can make in a specific amount of time. The attacker will not be able to rapid-fire guess passwords on an account using programs and parallel computing. An example is only allowing one login attempt every 5 seconds, which would slow down the attacker's operations drastically. Another solution that is commonly used is to lockout an account after a certain number of attempts. Although it can result in an inconvenience for the user, if the unlocking process or password reset procedure is secure, this can save accounts from being hacked. Another method that is

becoming more popular is multi-factor authentication or MFA, which will be discussed later in this paper.

A specific form of a brute force attack is a dictionary attack. This entails guessing the password using a dictionary of leaked passwords or an actual English dictionary. The attacker may also append commonly used numbers or characters to the end of the words to try and guess the password, such as “123” or “!”. This method is limited to the dictionary content and is not as popular due to its time-consuming nature and low performance.

Phishing is another common form of obtaining a user’s credentials by throwing bait for the user to catch. It can take the form of an email, an illegitimate website, or even malware. The general idea is for the victim to be tricked into entering or sending their information to the phisher who can then exploit it. Anti-phishing solutions fall into two categories: phishing prevention and phishing detection. Phishing prevention includes multi-factor authentication, but can be bothersome when implemented as another device is often needed, such as a cell phone to receive a code via SMS every time the user logs in. Phishing detection can be divided into user awareness and software detection. User awareness requires the user to be educated on phishing techniques and be wary when visiting new webpages. This is prone to human error and not a reliable way to detect phishing. Software detection can be either the traditional blacklist method, or automatic detection. Common automatic detection methods include phishing detection toolbars and detection methods that utilize artificial intelligence, more specifically, supervised learning classification algorithms. However, the machine learning algorithm does not perform well on large scale data sets and therefore, deep learning is used to increase performance. Deep learning also has drawbacks including the time necessary to train the algorithm. This study explains the various methods of anti-phishing solutions and shows the need for more research on deep learning algorithms for anti-phishing as it proves to have the best performance but there is room for improvement regarding training time and resources needed.

3. Methods of Password Security Evaluation

Many are familiar with the common specifications that organizations require when creating a password. This can look like:

- At least 8 characters
- At least 1 uppercase character
- At least 1 digit
- At least 1 special character

However, many experts, including the National Institute of Standards and Technology, have concluded that these requirements are actually not making passwords more secure and are simply a hassle for users. In theory, it makes sense that increasing the password length would increase security, this is not the case. Rather, a minimum character length creates a floor for the brute force guesses. Additionally, users will more often than not, create passwords with exactly the minimum amount of characters. This makes the set of possible character combinations even smaller for the attacker.

Not to mention that the possible passwords, realistically are not all of the character combinations. They are restricted to English words, have common patterns, and are subject to other requirements. When the attacker knows this, they can leverage this information to limit their search space and narrow down the possibilities, making it easier for them to hack accounts. This is why it is essential for other methods of password security evaluation to be employed to ensure secure accounts.

Dropbox has created a password strength estimator, called zxcvbn, in an effort to replace the common LUDS composition requirement that is not sufficient in rating password strength accurately. zxcvbn uses multiple sources to see how common a password is, such as leaked password data sets, common names from census data, and common words from Wikipedia pages. It calculates how many guesses it would take to crack a password by its rank in these sources. For example, if a password is the 100th most common entry from a source, zxcvbn would calculate that it would take 100 attempts to guess. The estimator also takes into account keyboard patterns, reversed sequences, repeating characters, dates, and can even recognize multi-patterned passwords, such as “qwerty123”. It can

distinguish between the keyboard pattern, “qwerty”, and the common numeric sequence, “123”. The depth of the zxcvbn algorithm is what makes it so accurate and powerful as a strength estimator. It does not simply look at the number of character types and overall password length like LUDS does, but actually takes into account password patterns and trends and uses real data to check against. It does have limitations, such as being unable to interpret or pattern match misspelled words or differentiate between common and uncommon single characters that do not match a pattern. This does not discredit its strength and is a dynamic tool that can increase the security of users.

Another, more recent development in password security evaluation uses deep learning, specifically recurrent neural networks, to estimate a password’s strength. Researchers from a university in Korea developed a deep learning model to evaluate password strength by predicting if a password had already been leaked or not by learning the features of leaked passwords. It also uses existing password strength evaluation scores to predict the strength of the password chosen by the user. Based on the results of this model, it performs better than LUDS and the zxcvbn algorithm. The deep learning model also has a time complexity of O(1), making it suitable for low performance environments.

4. Methods of Authentication

In today’s world, with advancing technology and means to hack accounts, using just a password for authentication is not enough to secure an account and verify the user. Other means are being used in conjunction with passwords to make sure the user logging in is authorized. Dual-factor or 2-step authentication is a security process which requires the user to provide two different methods of verification to ensure they are authorized. Usually this entails a password plus another form of identity verification. In this section, various methods of authentication will be discussed and explained.

Possession-based authentication requires the user to physically have an object to verify their identity. In other words, it is something only the user has. This is a form of authentication that everyone is familiar with. A few examples are keys, bank cards, and ID cards. This could be something like using another device or security token to sign in.

One type is a disconnected token which displays a randomly generated pin that changes every 30 seconds, for example, and is connected to the user's account. Everytime they sign in, they must provide the pin that is displayed on their token, in addition to their password. The disconnected token can also take the form of an app on the user's phone, instead of a physical token. Another type of token is a connected token. This is a device that needs to physically be plugged in or connected to the computer or primary device that the user is attempting to login to. Possession-based authentication is a basic but strong method, especially when used in dual-factor authentication, because the user must physically possess the object. Unlike passwords, which can be guessed or obtained through other various password attacks, it is harder to physically steal the object required from the user.

Another, newer form of authentication is biometrics. This entails utilizing "who you are" to authenticate an account. A few examples are a fingerprint, facial recognition, and voice recognition. It identifies the user based on a physiological trait that is unique to them. The users' biometrics cannot be reverse-engineered, in other words, no one can duplicate them. They are stored as binary digits using an encryption algorithm that does not allow you to reconvert them to images. Other benefits include the inability to misplace or forget a biometric feature and the difficulty of "faking" one. Despite these benefits, there are still system errors that can occur. Type I and type II errors (false negatives and false positives) can still occur. Therefore, it is best to use this authentication method in combination with other methods to provide the most security.

The last form of authentication that will be discussed in this paper is risk-based. This method uses an algorithm that estimates the risk of whether a login is a legitimate attempt from the user or an attack attempt by assessing various features and characteristics of the login such as time, location, IP address, device, etc. and calculating a risk score. It also uses previous login history and known attack data to analyze and estimate the risk. If the features of a login differ significantly from previous ones, the algorithm will request a re-authentication from the user for verification. A study was done on a large-scale online service using risk-based authentication and it was found that "RBA rarely requests

re-authentication in practice, even when blocking more than 99% of targeted attackers." The benefit of this is to not inconvenience users with having to dual-authenticate every time they log in and blocking attacks by using login features

5. Conclusion and Best Practices

In order to combat against password attacks, simple password requirements are not enough. Attackers are able to use more advanced technology to try and gain access to accounts and obtain potentially valuable information. Therefore, a combination of both password strength evaluation and login authentication methods are required to provide the best protection against attacks. Based on this paper's findings, it is not recommended for organizations to solely rely on the LUDS (lower-case, upper-case, digits, and symbols) requirement, due to the fact that users' tend to use real words or leetspeak instead of a secure combination of characters to create their passwords. Attackers can leverage this information to limit their search space when guessing passwords. Instead, these requirements could be combined with Dropbox's password strength estimator, zxcvbn. It is an algorithm that determines how secure a password is based on common passwords, leaked password data, various character patterns and common words. It performs well and is a much better indicator of password strength than LUDS. The first step in fighting against attackers is creating a secure password that is hard to hack.

As for login authentication, between the methods discussed in this paper, biometrics come out to be the most reliable way to identify the user. With technology advancing, it should not be a surprise to soon see most devices come with the option to use mechanics such as facial recognition or fingerprint identification. This is because a user's biometrics cannot be lost or forgotten and it is very difficult to replicate. However, it is not to say that the technology is perfect, so it must be combined with other forms of authentication to secure an account. Having multi-factor authentication to use in conjunction with biometrics would strengthen the security of accounts. For example, having biometrics be the first tier, due to its accuracy and ease for the user. Then, if there happens to be an error or identification failed, the next step could be a password, during which risk-based

authentication is used to ensure the login request is not suspicious. Which, if it is, possession-based authentication could be used, where the user must enter a pin sent to another device to be able to login. Having multiple layers of security authentication, combined with a strong password in the first place, would decrease the risk of attacks and is vital to keep data safe, ensuring only authorized users can gain access.

References

Farik, Mohammed & Lal, Nilesh & Prasad, Shalendra. (2016). A Review Of Authentication Methods. *International Journal of Scientific & Technology Research*. 5(11). 246-249.

Hawa Apandi, S., Sallim, J., & Mohd Sidek, R. (2020). Types of anti-phishing solutions for phishing attack. *IOP Conference Series: Materials Science and Engineering*, *769*(1), 012072. https://doi.org/10.1088/1757-899x/769/1/012072

Hong, K. H., & Lee, B. M. (2022). A deep learning-based password security evaluation model. *Applied Sciences*, *12*(5), 2404. https://doi.org/10.3390/app12052404

Hub, M., & Capek, J. (2011). Security evaluation of passwords used on internet. *Journal of Algorithms & Computational Technology*, *5*(3), 437–450. https://doi.org/10.1260/1748-3018.5.3.437

Papathanasaki, M., Maglaras, L., & Ayres, N. (2022). Modern Authentication Methods: A Comprehensive Survey. AI, *Computer Science and Robotics Technology*, 2022, 1–24. doi:10.5772/acrt.08

Wheeler, D. L. (1970, January 1). *Zxcvbn: {low-budget} password strength estimation*. USENIX. Retrieved October 15, 2022, from https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/wheeler

Wiefling, S., Jørgensen, P. R., Thunem, S., & Iacono, L. L. (2022). Pump up password security! evaluating and enhancing risk-based authentication on a real-world large-scale online service. *ACM Transactions on Privacy and Security*. https://doi.org/10.1145/3546069