Leveraging LLMs to protect users against target password guessing

Manaswi Raj[*]
manaswiraj195@kgpian.iitkgp.ac.in
Indian Institute of Technology
Kharagpur
Kharagpur, West Bengal, India

Harsh Sharma[*]
harshsharma2024@gmail.com
Indian Institute of Technology
Kharagpur
Kharagpur, West Bengal, India

Vibhor Dave[*]
vibhor.dave03@gmail.com
Indian Institute of Technology
Kharagpur
Kharagpur, West Bengal, India

Abstract

With the increasing sophistication of password-guessing attacks, traditional user-generated passwords have become more susceptible to compromise, particularly due to predictable patterns and common substitutions that attackers exploit. This study explores the use of Large Language Models (LLMs) to generate robust passwords that are resistant to targeted guessing attacks. By comparing LLM-generated passwords with user-created passwords, this research aims to determine if LLMs can produce secure, memorable passwords that address common vulnerabilities. Using an online survey, we will collect both user-generated and LLM-generated passwords from participants and analyze their resistance to advanced guessing algorithms. The study's findings seek to contribute to cybersecurity practices by demonstrating whether LLMs can provide a practical solution for enhancing password security, offering insights that may guide future developments in secure password creation methods.

ACM Reference Format:
Manaswi Raj, Harsh Sharma, and Vibhor Dave. 2018. Leveraging LLMs to protect users against target password guessing. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (USP '24).* ACM, New York, NY, USA, 4 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

In today's digital age, password-based authentication is still the primary security mechanism for most online platforms and services. However, user-generated passwords are often predictable and vulnerable to sophisticated password-guessing attacks. Attackers increasingly leverage machine learning (ML) and other data-driven methods to predict passwords based on common user tendencies, such as using easily memorable patterns or predictable substitutions (e.g., replacing "a" with "@"). Traditional password creation techniques fail to provide adequate protection against these advanced, targeted guessing attacks. This growing risk has highlighted the need for stronger, more unpredictable passwords that balance security with ease of recall.

Recent research has explored using Large Language Models (LLMs) for generating robust passwords that are resistant to such attacks [1]. Additionally, targeted online password guessing remains a significant security concern, as demonstrated by studies that examine the effectiveness of such attacks [2].

2 Research Questions

Our research questions were developed to address critical gaps in the understanding of LLM-generated passwords' effectiveness. They focus on two main aspects: the inherent guessability of LLM-generated passwords, and the influence of users' historical password data on the guessability of both LLM-generated and user-generated passwords.

Research Question 1

Does the LLM's password generation model, when tuned to avoid common user patterns, produce passwords that are statistically less guessable than user-created passwords derived from similar patterns?

Context: This question examines whether LLM-generated passwords, specifically tuned to avoid common user patterns, are statistically less guessable than passwords created by users. By comparing these two types of passwords, this study seeks to determine if LLMs can effectively minimize guessability by circumventing the predictable patterns users often employ.

Variables:

- **Password Type:** Categorical variable indicating if a password is LLM-generated (with tuning to avoid common patterns) or user-generated. This variable helps

[*] All authors contributed equally to this research.

USP '24, 10-09-24, Kharagpur, IN

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/XXXXXXX.XXXXXXX

differentiate and compare the guessability of each password source.

- **Guessability (targuess):** A continuous variable measuring the number of attempts needed by a targeted guessing algorithm to crack the password, representing password resilience.
- **Pattern Avoidance:** Binary variable indicating whether the password avoids common patterns (e.g., sequences, common substitutions). This variable is essential for examining the role of pattern avoidance in LLM-generated passwords.

Research Question 2

Does the number of previous passwords (*n*) provided by a user influence the guessability of an LLM-generated password, compared to the user's own subsequent password?

Context: This question investigates whether the quantity of prior passwords known to a targeted guessing algorithm impacts the security of LLM-generated passwords relative to user-generated ones. It aims to determine if a greater number of previous passwords affects the guessability of these two types of passwords differently, providing insight into how password history might impact targeted guessing effectiveness.

Variables:

- **Number of Previous Passwords (*n*):** A numerical variable representing the count of a user's prior passwords provided to the guessing algorithm, affecting its ability to predict subsequent passwords.
- **Password Type:** Categorical variable (LLM-generated or user-generated) used to compare the resilience of each password type.
- **Guessability (targuess):** A continuous variable that measures the number of attempts required to crack the password using targeted guessing methods.
- **Subsequent User Password:** Categorical variable for the user's password following their prior passwords, used for comparison against LLM-generated passwords to assess relative guessability.

3 Data Collection Architecture

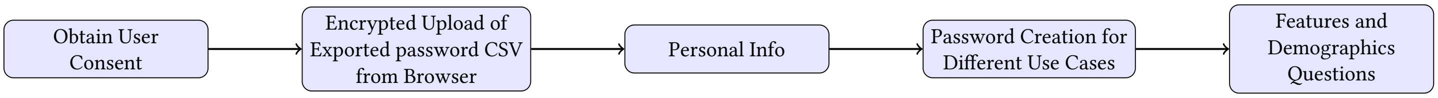
Obtain User Consent
Encrypted Upload of Exported password CSV from Browser
Personal Info
Password Creation for Different Use Cases
Features and Demographics Questions

Figure 1. Data Collection Architecture

The data collection architecture consists of the following steps:

1. **Obtain User Consent:** The process begins by obtaining consent from the user. This is a necessary step to ensure that the user agrees to share their information in compliance with privacy policies.
2. **Encrypted Upload of Exported Password CSV from Browser:** After consent is obtained, the user securely uploads an exported CSV file of their saved passwords from their browser. This file is encrypted during the upload to protect sensitive data.
3. **Personal Info:** The system then collects personal information related to the user. This information may include details necessary to tailor password recommendations or for demographic analysis.
4. **Password Creation for Different Use Cases:** Based on the personal information and user requirements, the system prompts the user to create secure passwords tailored for different use cases. The specific categories include:
   - **Social Media:** Passwords for social media accounts to ensure secure and private access.
   - **Bank Password:** Strong passwords for online banking and financial accounts to enhance security against unauthorized access.
   - **Work Related:** Passwords for work-related accounts or applications, designed to meet organizational security policies.
   - **Unimportant Sites:** Passwords for websites or applications that are not very important for users from privacy point of view.
5. **Features and Demographics Questions:** Finally, the system collects additional feature-related and demographic information. This data helps in further refining the password generation process and provides insights into user needs and preferences.

This architecture provides a comprehensive and secure framework for managing user data. By systematically obtaining consent, securely handling sensitive information, and tailoring password recommendations, it balances usability with privacy. Each step is designed to protect user data, reduce security risks, and enhance trust, making it a robust solution for responsible data collection and utilization.

4 Dataset Description

The dataset used in this study consists of user information relevant to understanding password usage patterns and demographics. This data has been carefully collected with user consent to ensure compliance with privacy standards and to provide meaningful insights for secure password recommendations. The dataset includes various attributes, each contributing to a comprehensive understanding of user profiles and security needs.

The dataset contains the following columns:

- **Full Name**: The full name of the user.
- **Mail**: The user's email address.
- **phone_number**: The user's phone number.
- **DOB**: The user's date of birth.
- **Gender**: The gender of the user.
- **Country**: The country where the user resides.
- **Previous Passwords**: A list of previous passwords used by the user, potentially collected from an uploaded CSV.

Example Data

Full Name	**Mail**	**phone_number**	**DOB**	**Gender**	**Country**	**Previous Passwords**
John Doe	john.doe@example.com	123-456-7890	1990-01-15	Male	USA	john123, qwerty123, doe@john
Jane Smith	jane.smith@example.com	987-654-3210	1985-05-23	Female	Canada	smith@123, janeSmith, SmithJane123
Alex Johnson	alex.j@example.com	555-234-5678	1992-12-01	Non-binary	UK	mypassword, password!2022, password@123

Table 1. Sample User Data with Passwords

5 Study Architecture

The study is designed to evaluate the effectiveness of LLM-generated passwords versus user-generated passwords by assessing their guessability, especially in cases where an attacker has prior knowledge of a user's password history. The study involves several stages, including data collection, password generation, and a comparative analysis using targeted guessing algorithms.

1. Data Collection

Participants will be asked to provide both personal and password-related information to simulate a realistic set of data that could be accessible to a potential attacker as mentioned in section 3.

2. Password Generation

After collecting personal and password-related data, we will generate two types of passwords for each participant:

- **User-Generated Password:** Participants will be asked to create a new password based on their typical habits and any specific context.
- **LLM-Generated Password:** Using the participant's personal information and previous password data, we will prompt an LLM to generate a password. The LLM will be fine-tuned to avoid common user patterns, leveraging the input data to produce a complex password that may or may not align with user tendencies.

3. Guessability Analysis using Targeted Guessing (targuess)

To assess the security of both types of passwords, we will use the ***targuess*** tool, which estimates the probability of successfully guessing a new password given previous password data. By analyzing both LLM-generated and user-generated passwords under *targuess*, we can compare the guessability scores for each password type, indicating which passwords are more resistant to targeted guessing attacks.

4. Statistical Analysis

After obtaining the guessability scores for both types of passwords, we will perform statistical analyses to evaluate the effectiveness of LLM-generated passwords compared to user-generated passwords. Some suggested statistical tests include:

- **Paired t-test:** To compare the average guessability scores (probability of guessing) of LLM-generated and user-generated passwords within the same participant group. This test will help identify if there is a statistically significant difference in guessability between the two types.
- **ANOVA (Analysis of Variance):** To evaluate if guessability scores vary significantly across different categories of passwords (e.g., social media, bank, work-related, forceful sites) when generated by both users

and LLMs. This test can identify if context influences password strength differently across the two methods.

- **Correlation Analysis:** To examine relationships between variables, such as the number of previous passwords (n) and the guessability score of the new password. This can help identify if knowledge of more previous passwords increases the guessability of user- or LLM-generated passwords.

5. Interpretation and Reporting

The results from the guessability analysis and statistical tests will be used to draw conclusions about the comparative effectiveness of LLM-generated and user-generated passwords. We will evaluate whether LLM-generated passwords offer significant security benefits and if they exhibit less predictability than user-created passwords based on similar data inputs. The findings will be compiled into a report that discusses the strengths and weaknesses of each approach and provides recommendations for enhancing password security using AI-generated methods.

6 Study Drawbacks

While this study offers valuable insights into the effectiveness of LLM-generated passwords compared to user-generated ones, several limitations should be considered:

1. **Dependence on User-Generated Data Quality:** The study relies on users providing accurate and honest information, especially regarding previous passwords and demographic details. Any inaccuracies in the data could skew the results, as the guessability of passwords generated by the model is highly dependent on historical patterns.
2. **Generalizability of Results:** The study's findings may be limited to the specific demographic and behavioral patterns of the participants. If the participant group lacks diversity, the results may not generalize well to a broader population with different cultural or security practices.
3. **Guessability Tool Limitations:** The effectiveness of the *targuess* tool in assessing password security depends on its ability to simulate realistic attack scenarios. If the tool does not accurately represent advanced password-guessing techniques, it may not provide a fully reliable measure of the generated passwords' resilience.
4. **Ethical and Privacy Concerns:** Collecting sensitive data, such as personal information and previous passwords, raises privacy and ethical concerns. Although steps are taken to secure this data, any vulnerability could lead to serious privacy breaches.
5. **Potential Overfitting to Known Patterns:** If the LLM is trained on commonly used password patterns, it may inadvertently reinforce predictable elements, limiting the uniqueness of generated passwords and their robustness against informed attackers.

References

[1] Rando, J., Perez-Cruz, F., and Hitaj, B. Passgpt: password modeling and (guided) generation with large language models. In *European Symposium on Research in Computer Security* (2023), Springer, pp. 164–183.

[2] Wang, D., Zhang, Z., Wang, P., Yan, J., and Huang, X. Targeted online password guessing: An underestimated threat. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security* (2016), pp. 1242–1254.