A Lightweight Markov-Model Approach to Password-Strength Assessment and Attack-Risk Prediction

徐芳澄（B11130108）　鄭兆翔（A11317005）　莊又翰（B11102021）
National Taiwan University of Science and Technology

May 21, 2025

Abstract

Passwords remain the most pervasive single-factor authentication method, yet users frequently adopt predictable patterns. This paper first surveys modern machine-learning techniques for password-strength estimation, then presents a lightweight 4-gram Markov model trained on a 70 000-sample subset of the RockYou leak. Log-likelihood scores produced by the model correlate with guessability and are compared against the popular `zxcvbn` meter. Receiver–operating-characteristic analysis shows the Markov approach identifies weak passwords with an area-under-curve of 0.82, supporting earlier findings that even modest statistical language models capture much of human password behaviour (Castelluccia et al., 2012). The open-source implementation provides educators with a computationally inexpensive tool for illustrating practical attack scenarios, while underscoring the continued importance of length-oriented policies and multifactor authentication.

1 Introduction

Text passwords are a convenient but fragile line of defence against unauthorised access. Traditional meters focus on length and character-class diversity, yet ignore structural patterns that make passwords predictable (Wheeler, 2016). Machine-learning (ML) models trained on large leaks can approximate an attacker's probability distribution, offering more realistic strength assessment (Melicher et al., 2016). This report (i) synthesises recent literature on ML-based password analysis and (ii) implements a 4-gram Markov model to demonstrate a fast, low-budget alternative to complex neural approaches.

2 Literature Review

2.1 Statistical Language Models

n-gram Markov Chains. Early work by Castelluccia et al. (2012) built adaptive strength meters using 3–5-gram models. Short-order chains balance coverage and data sparsity but cannot capture

long-range structure.

Probabilistic Context-Free Grammars (PCFGs). Weir et al. (2009) introduced PCFGs to model high-level structures such as *dictionary-word + digits*. Subsequent variants transfer learned grammars across length regimes (Han et al., 2021), yet grammar induction remains labour-intensive.

2.2 Neural Approaches

Recurrent Neural Networks. RNN-based language models learn longer-range dependencies and achieve state-of-the-art cracking efficiency with moderate hardware (Melicher et al., 2016), though training is slow and interpretability limited.

Generative Adversarial Networks. PassGAN learns password distributions without handcrafted rules (Hitaj et al., 2019), while DenseGAN extends this with DenseNet blocks (Fu et al., 2021). Despite their power, GANs are notoriously hard to stabilise.

2.3 Expert Systems

`zxcvbn` combines large dictionaries, keyboard-walk detection and entropy estimation, remaining the industry baseline for user-facing meters (Wheeler, 2016). Hybrid designs that fuse ML probabilities with `zxcvbn` pattern matching appear most robust (Liu et al., 2018; Ur et al., 2017).

3 Methodology

3.1 Dataset

Seventy-thousand unique ASCII passwords were randomly sampled from the RockYou leak, lowercased and trimmed.

3.2 Four-Gram Markov Model

Given a password $w = c_1 \dots c_m$, the model estimates $P(c_{i+3} \mid c_i c_{i+1} c_{i+2})$ for every contiguous 4-gram, using Laplace smoothing ($k = 1$). The cumulative log-likelihood

$$\log L(w) = \sum_{i=1}^{m-3} \log P(c_{i+3} \mid c_i c_{i+1} c_{i+2})$$

serves as the strength score.

3.3 Baseline and Metrics

Passwords receive the label *weak* when `zxcvbn` outputs a score < 3; otherwise *strong*. We compute ROC curves, AUC, precision–recall and approximate guess rank by ordering passwords ascending in $\log L$.

4 Results and Discussion

The Markov model attains an AUC of 0.82 against `zxcvbn`. It agrees on 79 % of `zxcvbn`'s weak classifications but diverges on long keyboard walks—high Markov probability yet score 0—and on obscure multi-word passphrases absent from RockYou—low Markov probability, moderate-to-high `zxcvbn` score. Complexity rules that encourage symbol substitution raise `zxcvbn` scores more than they lower $\log L$, echoing the findings of Ur et al. (2017).

5 Conclusion and Future Work

Even lightweight statistical learning provides actionable insight into password weakness. Educators can embed our open-source code into lab sessions to visualise real attack economics. Future work will explore Transformer language models and incremental retraining on new leaks to reduce dataset bias.

Team Contribution

組員（學號）	主要工作內容	備註
徐芳澄（B11130108）	文獻蒐集與撰寫，簡報設計	負責 APA 引用格式
鄭兆翔（A11317005）	Demo 開發（4-gram Markov），實驗與結果撰寫，整合全文	報告第一作者
莊又翰（B11102021）	資料清理，ROC/AUC 腳本，圖表製作，校對排版	交叉檢查數值

References

Castelluccia, C., Dürmuth, M., and Perito, D. (2012). Adaptive password-strength meters from markov models. In *Proceedings of NDSS*.

Fu, C., Duan, M., Dai, X., Wei, Q., Wu, Q., and Zhou, R. (2021). Densegan: A password guessing model based on densenet and passgan. In *Information Security Practice and Experience (ISPEC)*, volume 13025 of *Lecture Notes in Computer Science*, pages 296–305. Springer.

Han, W., Xu, M., Zhang, J., Wang, C., Zhang, K., and Wang, X. (2021). Transpcfg: Transferring the grammars from short passwords to guess long passwords effectively. *IEEE Transactions on Information Forensics and Security*, 16:451–465.

Hitaj, B., Gasti, P., Ateniese, G., and Pérez-Cruz, F. (2019). Passgan: A deep learning approach for password guessing. In *Applied Cryptography and Network Security*, volume 11464 of *Lecture Notes in Computer Science*, pages 217–237. Springer.

Liu, Y., Xia, Z., Yi, P., Yao, Y., Xie, T., Wang, W., and Zhu, T. (2018). Genpass: A general deep learning model for password guessing with pcfg rules and adversarial generation. In *Proceedings of IEEE ICC*, pages 1–6. IEEE.

Melicher, W., Ur, B., Segreti, S. M., Komanduri, S., Bauer, L., Christin, N., and Cranor, L. F. (2016). Fast, lean, and accurate: Modeling password guessability using neural networks. In *Proceedings of the 25th USENIX Security Symposium*, pages 175–191.

Ur, B., Segreti, S. M., Bauer, L., Christin, N., Cranor, L. F., Melicher, W., and Roberts, H. (2017). Design and evaluation of a data-driven password meter. In *2017 IEEE Symposium on Security and Privacy*, pages 524–541.

Weir, M., Aggarwal, S., de Medeiros, B., and Glodek, B. (2009). Password cracking using probabilistic context-free grammars. In *Proceedings of the 30th IEEE Symposium on Security and Privacy*, pages 391–405. IEEE.

Wheeler, D. L. (2016). zxcvbn: Low-budget password strength estimation. In *Proceedings of the 25th USENIX Security Symposium*, pages 157–173.