

人間計算可能なパスワードに対する 畳み込みニューラルネットワークを用いた安全性評価

小川翔平, 櫻井幸一
(九州大学大学院システム情報科学府)

1. はじめに

近年、様々な情報サービスにおいて、パスワード認証は広く利用されているが、パスワードの使い回しによるセキュリティ上のリスクが懸念されている。このような問題に対して、「人間計算可能なパスワード」という新しい認証方式が提案された。

本研究では、この方式の安全性を機械学習の観点から検証し、どの程度攻撃が成立しうるかを評価するために、新たなモデルを用いて予測精度の限界を探った。

2. 既存研究

2.1. 人間計算可能なパスワード

人間計算可能なパスワードとは、2016年にManuel Blumらによって提案された認証方式であり、コンピュータを用いることなく、ユーザが自身の脳内で計算を完了させるチャレンジレスポンス認証方式の一種である。

この方式は、ユーザがパスワードを生成・記憶する段階である「PRPE」と、実際に認証を行う段階「PROC」で構成される。PRPEは更に、秘密鍵の記憶 (MEM) と計算方法の習得 (COM) を行う工程に分割される。PROCは、ユーザが記憶した秘密鍵と提示されたチャレンジを用いて計算を行い、認証応答を生成する段階である。これらのプロセスにより、認証ごとに毎回異なるパスワードを生成できるのが特徴である。

| 画像番号 | 0 | 1 | 2 | 3 | ... | 10 | 11 | 12 | 13 |
|-------------|---|---|---|---|-----|----|----|----|----|
| 画像 I | | | | | ... | | | | |
| $\sigma(I)$ | 1 | 2 | 8 | 6 | ... | 4 | 9 | 7 | 3 |

$$j = \sigma(\text{Image 10}) + \sigma(\text{Image 11}) \bmod 10 = 3$$

$$f(x_0, x_1, \dots, x_{13}) = \sigma(\text{Image 0}) + \sigma(\text{Image 1}) + \sigma(\text{Image 2}) \bmod 10 = 6$$

図1. ユーザが記憶する関数群とその計算例

2.2. 双方向長短期記憶による安全性評価

既存研究では、攻撃手法の一つとして、埋め込み層を前処理として導入した双方向長短期記憶を用いた解析が実施されている。この研究では、攻撃者が通信経路などを介してユーザの認証情報の一部を取得した状況を想定し、機械学習を用いたなりすまし攻撃の可能性を検証している。具体的には、14枚のチャレンジ画像と、入力したレスポンスのデータセットが漏洩したケースを考える。

結果として、ユーザが記憶すべき画像枚数が26枚の場合、予測精度は最大で55.81%に達することが確認された。一方、画像枚数が50枚および100枚の場合の予測精度は最大で17.20%にとどまり、ユーザが記憶する画像枚数が増加するにつれて予測が困難になることが示されている。

3. 研究手法

本研究では、埋め込み層を導入した畳み込みニューラルネットワーク (CNN) を使用することで、認証情報の予測精度向上の可能性を検討した。具体的には、それぞれ次の式によって定義される人間計算可能関数を対象と

している。

$$f_{2,2}(x_0, \dots, x_{13}) = x_j + x_{12} + x_{13} \bmod 10$$

ただし、 $j = x_{10} + x_{11} \bmod 10$

$$f_{1,3}(x_0, \dots, x_{13}) = x_j + x_{11} + x_{12} + x_{13} \bmod 10$$

ただし、 $j = x_{10} \bmod 10$

4. 結果

ユーザが記憶する画像枚数 $N=26$ の場合、データセット数50000において予測精度は最大で60.84%に達した。一方で、その他の場合の予測精度は最大でも約20%にとどまり、ランダムに予測した場合とほとんど変わらない結果となった。特に、画像枚数を $N=50$ または $N=100$ に設定した場合、データセット数が50000に達すると、訓練データに対する学習すら進まない現象が確認された。また、ユーザが記憶する画像枚数 N が増加するにつれて、予測精度が低下する傾向が見られた。

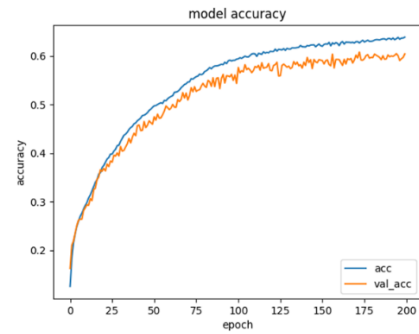


図2. CNNによる $f_{2,2}$ の予測精度 (画像枚数: $N=26$)

5. まとめ

既存研究のMLPやRNN系モデルに比べ、CNNを導入することで秘密関数推測の精度向上が可能であることが示唆された。人間計算可能なパスワードは高い安全性を持つが、一定の計算パターンが存在するため、チャレンジレスポンスデータが漏洩した場合の攻撃リスクは否定できない。

参考文献

- [1] Manuel Blum, Santosh Vempala. The complexity of human computation via a concrete model with an application to passwords. Proceedings of the National Academy of Sciences. July 2017.
- [2] Murata, I., He, P., Gu, Y., Sakurai, K. (2023). Towards Evaluating the Security of Human Computable Passwords Using Neural Networks. WISA 2022. Lecture Notes in Computer Science, vol 13720. Springer.
- [3] 丸野美矩人, 人間計算可能なパスワードに対する長短期記憶ニューラルネットワークによる安全性評価, 九州大学工学部電気情報工学科卒業論文, 2023.
- [4] 池田 悠登, 櫻井 幸一. 人間計算可能なパスワード認証に対する埋め込み型の双方向LSTMを用いた安全性実験評価. IPSJ/CSEC2024-63.