"Enhancing Mobile Authentication Security: A Machine-Learning Approach with Jumbled Number Pad Two-digit PIN Analysis"

Rishikar Chittimalla

Abstract

This project addresses the increasing need for enhanced security in mobile authentication systems, particularly for Android devices. Traditional authentication mechanisms often compromise on either security or user convenience. To address this, we propose a novel authentication system that leverages a jumbled number pad combined with machine learning to analyse users' behavioural patterns during authentication. The system requires users to remember a six-digit password along with a two-digit code. The authentication process involves the user entering their password on a randomly jumbled number pad, with the machine learning model analysing the timing and pattern of key entries (especially, those two-digit code) to authenticate the user. Preliminary studies indicate a significant improvement in security against common attack vectors such as shoulder surfing and brute force attacks, while also offering an intuitive user experience. This study contributes to the field by providing a novel approach to secure mobile authentication that balances robust security measures with user convenience.

Motivation

The motivation behind this project stems from the need to address the limitations of existing mobile authentication methods. For instance, fixed-pattern PINs are susceptible to shoulder surfing, while biometrics can be compromised and are not always reliable under varying conditions. By leveraging machine learning to understand and authenticate based on user behavior, this system aims to introduce an additional layer of security that is difficult to replicate or breach by malicious entities.

Introduction

In an era where mobile devices have become ubiquitous, securing sensitive information against unauthorized access is paramount. Traditional authentication methods, including PINs, patterns, and biometrics, have been widely adopted but each comes with its own set of vulnerabilities and inconveniences. This project introduces a novel authentication system designed to enhance security and user convenience by integrating a jumbled number pad with a machine-learning algorithm that analyses user-specific behavioural patterns during password entry.

Data Collection

1. Data is collected during user sessions where the participant enters their full six-digit password, including the two-digit code, on a jumbled number pad. Each session is recorded with timestamped keypress events.
2. Special attention is given to the entry of the two-digit code within the full password. This segment is crucial as it represents a unique behavioural pattern due to its significance to the user.

Feature Extraction

From the raw data, several features are extracted to capture the user's behavioural pattern:

1. The time taken for each key press, especially focusing on the two-digit code. This can indicate how familiar or hesitant the user is with the current jumbled layout.
2. The time interval between releasing one key and pressing the next, particularly between the first and second digits of the code, and between the code and adjacent digits in the password. This can provide insights into the user's flow and rhythm.
3. The overall speed of entering the two-digit code compared to the rest of the password. A significant deviation might indicate a unique user behaviour.
4. The frequency of corrections or mistypes when entering the two-digit code, which may reflect the user's comfort and accuracy with the jumbled keypad.

Model Selection

The selection of an ML model is critical and depends on the nature and dimensionality of the extracted features:

1. Sequential Models: Given the temporal nature of typing behaviour, sequential models like LSTM (Long Short-Term Memory) or GRU (Gated Recurrent Units) can effectively capture the dynamics of typing patterns over time.
2. Ensemble Methods: Combining multiple models, such as decision trees with sequential models, might improve performance by capturing both the temporal dynamics and the intricate decision boundaries of user behaviours.

Training Strategy

1. Supervised Learning: The model will be trained in a supervised manner, with labelled data indicating whether the entry attempt was legitimate or not.
2. Balanced Dataset: Care should be taken to ensure the dataset includes a balanced representation of legitimate and illegitimate entry attempts to prevent model bias.
3. Cross-Validation: Employ cross-validation techniques to evaluate the model's performance reliably, ensuring it generalizes well to unseen data.
4. Continuous Learning: Consider implementing a mechanism for the model to learn continuously from new data, allowing it to adapt to slight changes in user behaviour over time.