Enhancing Web Application Security with Machine Learning-Powered Device Fingerprinting

Abstract

Device fingerprinting offers a powerful mechanism to bolster web application security by passively identifying and tracking individual devices. This project proposes a novel system that leverages machine learning techniques to enhance the accuracy and robustness of device fingerprinting within the context of web applications. By unobtrusively collecting device attributes through JavaScript and applying machine learning algorithms, reliable device identification and anomaly detection can be achieved, improving security posture.

1 Introduction

1.1 Problem Statement

Traditional user authentication methods (e.g., usernames and passwords) are vulnerable to various attacks. Device fingerprinting provides a complementary layer of security by creating a unique identifier based on a device's hardware and software characteristics.

1.2 Research Gap

Existing fingerprinting techniques often lack adaptability and robustness against evolving device configurations.

1.3 Objectives

- Design and implement a JavaScript-based data collection module to gather comprehensive device attributes.
- Build machine learning models to refine device fingerprinting accuracy and reduce false positives.
- Develop an AI-based anomaly detection system to flag suspicious changes in device fingerprints, potentially indicating malicious activity.

2 Methodology

2.1 Comprehensive Data Collection

- **JavaScript Implementation:** A script will be embedded within web applications to collect:
  - WebGL Properties: GPU information for detailed device profiling.
  - Canvas Fingerprinting: Unique rendering characteristics for robust identification.
  - Browser Features: User-agent, screen resolution, plugins, etc.
  - Font Detection: Presence of installed fonts for fine-grained discrimination.
  - CPU and Hardware Specifications: Enhance fingerprint detail (if accessible).
  - Behavioral patterns: Scrolling habits, mouse movement, typing patterns (with careful ethical considerations) to strengthen the fingerprint.
- **Machine Learning Integration**
  - **Model Selection:** Exploration of algorithms suited for device fingerprinting, including:
    - Supervised approaches (e.g., Random Forests, Support Vector Machines) for classification.
    - Unsupervised techniques (e.g., clustering, isolation forests) for anomaly detection.
  - **Feature Engineering:** Careful selection of the most discriminating attributes and potential transformation of features to improve model performance.
  - **Continuous Learning:** The system will adapt to legitimate configuration changes while maintaining detection sensitivity.

2.2 Security and Privacy Considerations

- **Transparency:** Clear and concise communication to users about the purpose and methods of data collection.
- **Data Minimization:** Collect only necessary attributes for fingerprinting.
- **Secure Storage and Handling:** Implement robust encryption and access controls.
- **Compliance:** Adherence to relevant data privacy regulations (e.g., GDPR).

3 Expected Outcomes and Impact

- **Enhanced Security:** Strengthen user identification beyond conventional authentication methods, offering a persistent security mechanism.
- **Fraud Prevention:** Deter fraudulent activities by making it harder to spoof or impersonate legitimate devices.
- **Improved User Experience:** Potential reduction of friction in authentication processes for known and trusted devices.

4 Literature Review

- **Survey on Device Fingerprinting Techniques:** `https://arxiv.org/abs/1905.01051`
- **Machine Learning for Cybersecurity:** `https://www.cisco.com/c/en/us/products/security/machine-learning-security.html`