**CYBER SECURITY ASSIGNMENT-2**

**REPORT**

**Name:** Vemuri Sai Soumya

**Date:** 05-10-2025

**Course:** Cyber Security (22CIE55)

**GitHub Repository:** https://github.com/Soumyaa2005/CyberSecurity-Assignment-2

**Research Paper:** "You Again": Fingerprinting and Tracking Mechanisms of Malicious Sites

**Author:** Erin Kuffel-Flato

**ABSTRACT**

This report presents an analytical summary and extended study of the paper "You Again": Fingerprinting and Tracking Mechanisms of Malicious Sites by Erin Kuffel-Flato. The study explores how malicious websites use tracking and fingerprinting to monitor user activity, comparing these behaviors against legitimate websites. The aim of this report is to understand, analyze, and propose improvements to the methodologies discussed in the research, contributing to enhanced cybersecurity detection models.

**METHODOLOGY AND APPROACH**

The original research employed OpenWPM, an automated browser management platform, to analyze over 12,000 malicious URLs and 3,442 benign Tranco homepages. Six mechanisms were studied across stateful and stateless tracking methods, namely:  
- Stateful Tracking: HTTP Cookies, ETags, and HTML5 LocalStorage.  
- Stateless Fingerprinting: Canvas, WebRTC, and Navigator-based fingerprinting.  
  
The analysis revealed that while tracking is prevalent on both categories of websites, the diversity of tracking techniques is lower on malicious sites. Navigator querying and third-party dependencies remain dominant across the board. Interestingly, malicious sites mimic legitimate tracking behavior, making detection difficult using traditional binary identification approaches.

**KEY FINDINGS AND RESULTS**

- Tracking mechanisms are present in 81% of malicious and 94% of benign sites.  
- Navigator-based fingerprinting was detected on 53% of malicious and 73% of benign websites.  
- Third-party tracking scripts accounted for more than 90% of observed mechanisms.  
- Malicious first-party scripts focus heavily on Navigator fingerprinting (66.4%) and LocalStorage usage (78.2%).  
  
The similarity in behavioral composition between legitimate and malicious websites highlights the need for a deeper context-based approach to differentiate harmful scripts from functional tracking.

**IDENTIFIED RESEARCH GAP**

The primary gap identified lies in the inability of current methods to distinguish malicious from benign first-party tracking behavior. The research suggested that deeper examination of properties queried through the window.navigator API could yield more meaningful indicators of intent. This report extends that suggestion by proposing a property-level entropy analysis model capable of quantitatively assessing fingerprinting aggressiveness.

**PROPOSED MODEL AND IMPROVEMENTS**

To address the research gap, a Granular Navigator Property Analysis (GNPA) model is proposed. This model introduces the concept of a Fingerprinting Entropy Score (FES), which evaluates each script based on the uniqueness and persistence of the properties it queries.

|  |  |  |
| --- | --- | --- |
| Property Category | Example Properties | Associated Risk (Weight) |
| Low Entropy | userAgent, language, platform | 1 (Necessary Functionality) |
| High Entropy | hardwareConcurrency, deviceMemory, plugins, mimeTypes, getBattery() | 5 (Aggressive Fingerprinting) |

Each property’s contribution to the FES determines the risk level. A baseline calibration is performed using benign site data, and any first-party script exceeding the 95th percentile of benign FES scores is classified as malicious. This statistical threshold reduces false positives and enables context-aware malicious activity detection.

**CONCLUSION**

This enhanced approach refines the understanding of tracking mechanisms by focusing on Navigator property-level entropy and intent classification. The proposed GNPA model allows for precise differentiation between functional and aggressive fingerprinting, providing a foundation for improved browser-based detection tools. By integrating this methodology, cybersecurity systems can better identify and mitigate stealthy malicious tracking without hindering legitimate web functionalities.