Summary of Browser Fingerprinting Project

Vafa Andalibi

January 20, 2019

A browser fingerprint is a collection of various characteristics of a browser that form a unique identifier. Using a browser fingerprint, users can be tracked across different websites, which although the identification of the user might not be revealed, it is still a violation of users' privacy.

To defend against this violation, many users started using browser extensions to partially spoof their browser's attributes, e.g. user agent, etc. These extensions, however, might form an impossible set of configurations, e.g. a browser that claims to be an iPad but supports flash. This *Paradoxical* configuration makes the user even more visible in the sea of all users. Accordingly, this act of spoofing, needs to become more sophisticated.

In an attempt to address this issue, a Markov model was built on a partial dataset of 16000 browser fingerprints from AmIUnique project. This using model of browser fingerprint, *legitimate* browser attributes should theoretically be generated by just starting from the model and walking on Markov model nodes. Moreover, a set of browser attributes that is paradoxically spoofed should be revealed by applying the model on it.

A tool called ParadoxCatcher[1] was built for this purpose. The paper submitted to PoPETs journal was rejected, mostly because the reviewers was mentioning that more empirical work should be done on this work.

To start an empirical study based on ParadoxCatcher, attempts are being made to create a big dataset of fingerprints that does not include spoofed attributes. We have also already been in contact to possibly use this in the Tor browser to improve the usability and functionality since Tor browser is currently unifying all of its instances to keep their fingerprint the same. They seem to be willing to randomize some of the attributes in case there is a good solution for it.

Given a big reliable dataset which can be generated using virtualization, we could have a reliable Markov model which can be leveraged to improve functionality as follows:
Suppose a Tor user wants to watch a video with the actual monitor's resolution which is Full HD, but without revealing a trackable or paradoxical fingerprint which can safely be changed in next visit. Using the reliable Markov

[1] https://paradoxcatcher.readthedocs.io

model, the user can start "generating" a browser characteristic by starting on the node "resolution=FullHd" and walking over the markov model.

The "fingerprint generator" could be a tool that many people might also use for different purposes, since most of the available fingerprint datasets are not public and almost all of them contains records of browsers that partially spoofed attributes.