

Publication

A Machine Learning Approach to Detect Differential Treatment of Anonymous Users

ConferencePaper (Artikel, die in Tagungsbänden erschienen sind)**ID** 4649503**Author(s)** Wagner, Isabel**Author(s) at UniBasel** [Wagner, Isabel](#) ;**Year** 2022**Title** A Machine Learning Approach to Detect Differential Treatment of Anonymous Users**Editor(s)** Atluri, Vijayalakshmi and Di Pietro, Roberto and Jensen, Christian D. and Meng, Weizhi**Book title (Conference Proceedings)** Computer Security - ESORICS 2022**Place of Conference** Copenhagen, Denmark**Year of Conference** 2022**Publisher** Springer Nature Switzerland**Pages** 3-23

Anonymous Internet use is essential to safeguard against mass surveillance and to protect privacy online. Tor Browser and the Tor anonymity network provide an effective and convenient way to browse the Internet anonymously. However, many websites make browsing inconvenient for anonymous users either by blocking access entirely, by blocking access to some functionality, or by using CAPTCHAs to make access more difficult. Prior work has relied on heuristics to study the extent to which anonymous users are treated differently. However, these heuristics either underestimated blocking or required extensive manual labeling. To address these shortcomings, here we propose a machine learning approach to detect when anonymous users are treated differently. We train binary and multi-class classifiers based on six feature sets and find that they perform very well on our test data (F1 scores 100%–94% for binary and 97%–84% for multi-class classifiers). Applying these classifiers to data collected from browsing 1,000 sites, including visits to subsites and executing search functionality, via 100 Tor exit nodes we find that 16.7% of landing pages inconvenience anonymous users, compared with 15.2% of subsites and 3.8% of search result pages. In particular, we find that websites hosted by Akamai, EdgeCast, and Cloudflare have significantly higher block rates than average, and that blocking of search results is dominated by Google which displays a block page or CAPTCHA for 39.8% of search result pages.

Series title Lecture Notes in Computer Science**Digital Object Identifier DOI** 10.1007/978-3-031-17146-8_1**Document type (ISI)** inproceedings