



MEASURING THE ACCESSIBILITY OF POPULAR WEBSITES WHILE USING MULLVAD VPN

Francine Biazin do Nascimento - supervised by Dr. Stefanie Roos



1 BACKGROUND

- Users who seek anonymity online use anonymity networks such as **VPNs**, **Tor**, **I2P**, etc.
- **Anonymity networks** work by sharing IP addresses among a pool of users
- Some shared IP addresses have been **blacklisted**
- Some web servers and content distribution networks (CDNs) **block** blacklisted IP addresses
- Users then experience excessive **CAPTCHAs**, **block pages**, etc.
- This type of blocking constitutes **server-side blocking**

2 RESEARCH QUESTIONS

- To what **extent** do websites block users accessing them through **Mullvad VPN**?
- What is the **nature** of these blocks?

3 METHOD

- Domains taken from **Alexa Top 10K Sites**
- **Crawl** websites from a **Mullvad VPN** connection and from a control connection (cf. Figure 1)
- Divide experiment into stages:
 - **Stages 0-2:** testing implementation and scalability
 - **Stage 3:** 3,000 domains, only home pages, Swedish exit node, Dutch control connection
 - **Stage 4:** 1,000 domains, 2 subpages from each, Dutch exit node, Dutch control connection
- **Set up:** two machines, two Internet connections - running in parallel

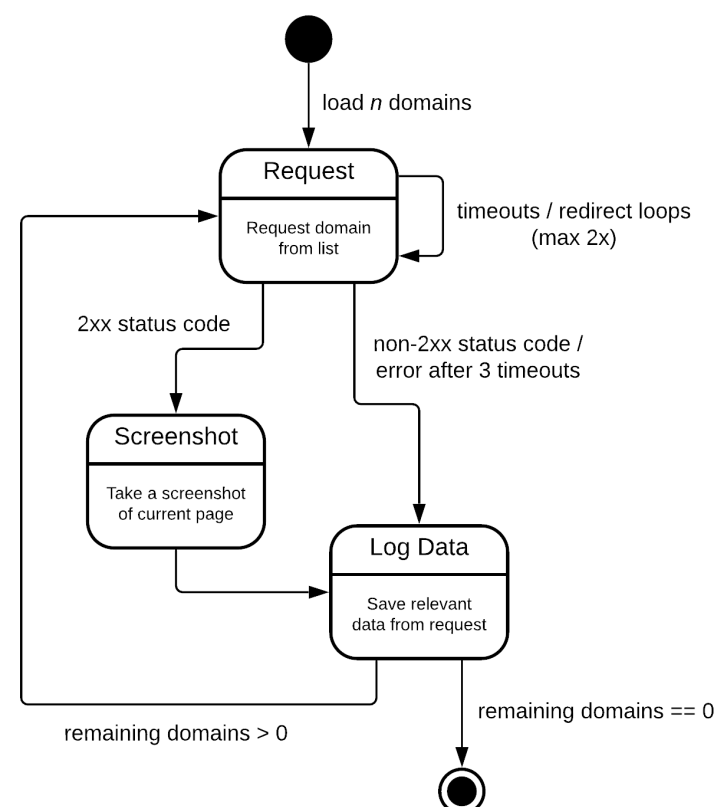


Fig. 1 State diagram for crawler.

4 BLOCK CLASSIFICATION

- Compare results from VPN connection to control
- **Not Blocked:** VPN connection return **200** status code
 - control also 200 and **screenshot comparison** indicate no blocking (perceptual hashing)
 - control **failed** (manual check)
- **Blocked:** control connection return **200** status code
 - VPN **failed**
 - **screenshot comparison** indicates blocking
- **No Difference:** both VPN and control connections failed with the same kind of response
- **Maybe Blocked:** both VPN and control connections failed, but with different kinds of responses
- For Stage 4, classification at **domain level**:
 - **Home Page Blocked:** could not access home page; domain classified as blocked
 - **Subpage Blocked:** could access home page, but one or more subpages blocked; domain classified as blocked

5 RESULTS

- **Stage 3:**
 - Blocked requests: **0.97%** (1.2% if 'Maybe Blocked' counted as 'Blocked')
 - Most common blocks: **HTTP blocks**, timeout blocks; **differentiated content** for successful requests
- **Stage 4:**
 - Individual requests: **0.75%** blocked - mostly **HTTP blocks**, followed by differentiated content
 - Domain level: **1.38%** blocked - mostly at **subpage** level
- **Categories:**
 - Used **McAfee URL categorisation service**
 - Results for Stage 3 show high blocking ratio for **'Restaurants'** (cf. Figure 2)
 - Results for Stage 4 show high blocking ratio for **'Remote Access'** (cf. Figure 3)

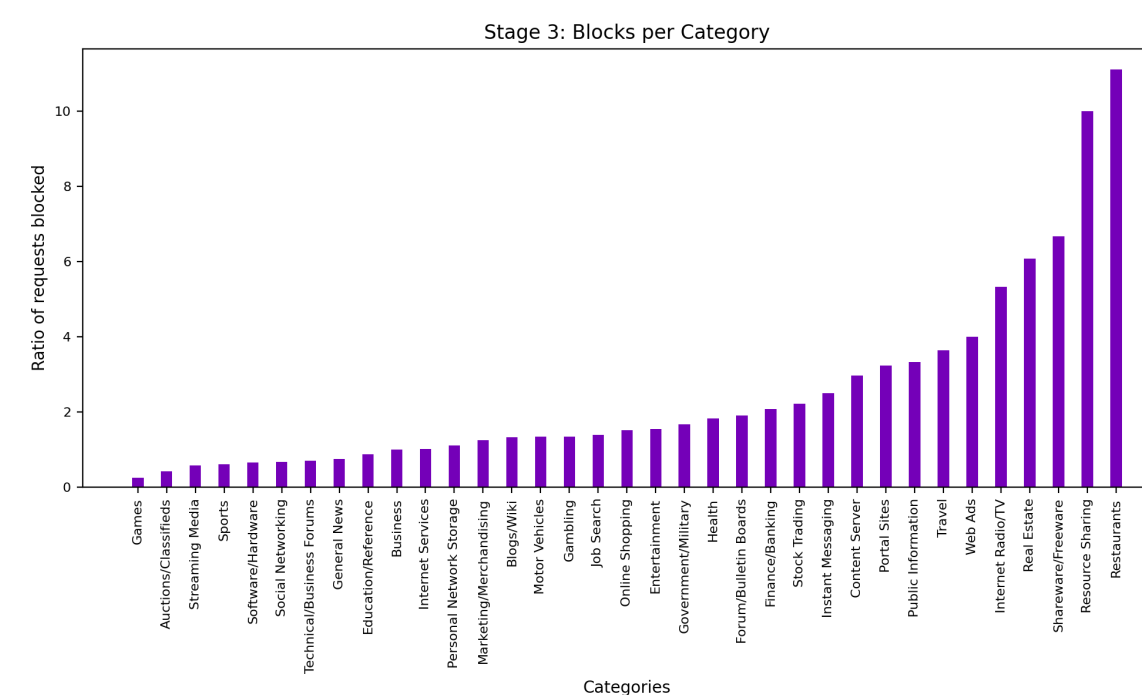


Fig. 2 Graph illustrating the ratio of blocked requests identified per category of website during Stage 3 (requesting home pages only).

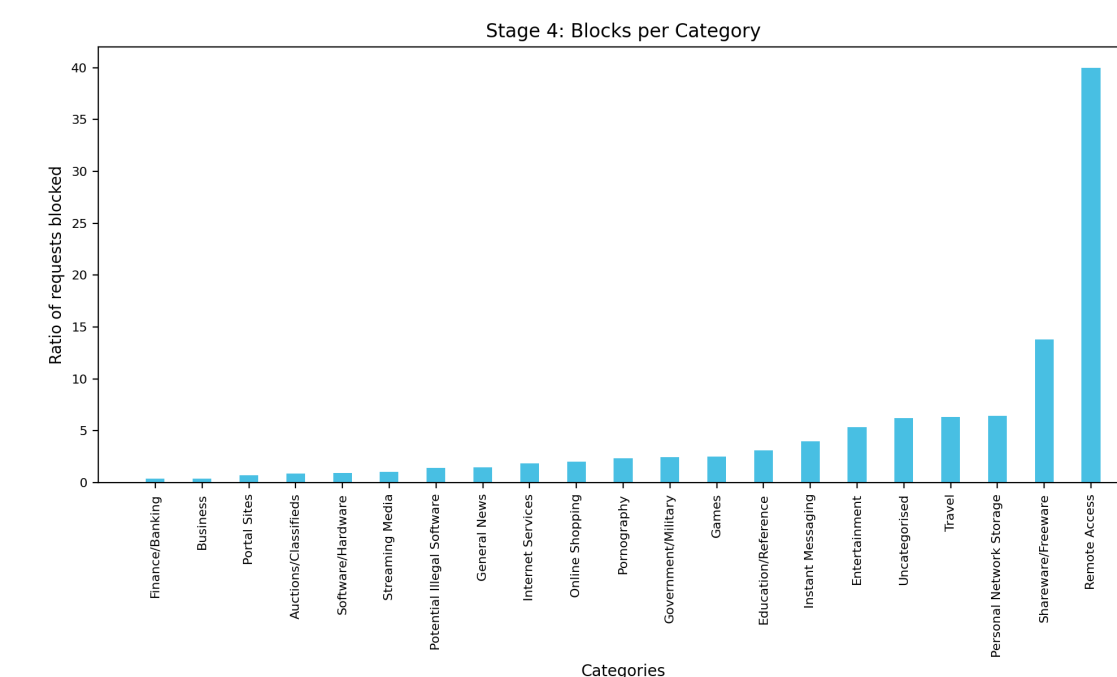


Fig. 3 Graph illustrating the ratio of blocked domains identified per category of website during Stage 4 (requesting two subpages from each of 1,000 domains).

6 CONCLUSIONS

- No statistical significance when looking at individual requests (Stages 3 and 4)
 - Compared against failures observed in control connection
- However, there is a difference when looking at **domain level** (Two-Sample Proportion test, p -value = $1.334e-18$, $\alpha = 0.025$)
- Deterioration in service also observed in categories with high blocking ratios, such as 'Health' and 'Government/Military'
- More discerning alternatives for user authentication using, e.g., zero-knowledge proofs

7 CONTACT

- Francine Biazin do Nascimento
- f.biazindonascimento@student.tudelft.nl